PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12Q 1/68, C07K 14/40	A2	(11) International Publication Number: WO 00/68420 (43) International Publication Date: 16 November 2000 (16.11.00)

(21) International Application Number: PCT/CA00/00533

(22) International Filing Date: 5 May 2000 (05.05.00)

(30) Priority Data:
60/132,878 5 May 1999 (05.05.99) US

(71) Applicant *(for all designated States except US)*: MYCOTA BIOSCIENCES INC. [CA/CA]; Suite 2550, 225, President-Kennedy West, Montreal, Quebec H2X 3Y8 (CA).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: ROEMER, Terry [CA/CA]; 4935 Connaught Avenue, Montreal, Quebec H4V 1X4 (CA). BUSSEY, Howard [CA/CA]; 325 Victoria Ave, Westmount, Quebec H3Z 2N1 (CA). DAVISON, John [CA/CA]; 4886 DeBullion, Montreal, Quebec H2T 1Z5 (CA).

(74) Agents: DUBUC, Jean, H. et al.; Goudreau Gage Dubuc, The Stock Exchange Tower, Suite 3400, 800 Place Victoria, Montreal, Quebec H4Z 1E9 (CA).

(81) Designated States: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: IDENTIFICATION OF *CANDIDA ALBICANS* ESSENTIAL FUNGAL SPECIFIC GENES AND USE THEREOF IN ANTIFUNGAL DRUG DISCOVERY

(57) Abstract

The invention relates to the identification and disruption of essential fungal specific genes isolated in the yeast pathogen *Candida albicans* namely *CaKRE5, CaALR1* and *CaCDC24* and to the use thereof in antifungal diagnosis and as essential antifungal targets in a fungal species for antifungal drug discovery. More specifically, the invention relates to the *CaKRE5, CaALR1* and *CaCDC24* genes, to their use to screen for antifungal compounds and to the drugs identified by such.

ATTORNEY DOCKET NUMBER: 10182-016-999
SERIAL NUMBER: 10/032,585
REFERENCE: CC

Description

TITLE OF THE INVENTION

IDENTIFICATION OF *CANDIDA ALBICANS* ESSENTIAL FUNGAL SPECIFIC GENES AND USE THEREOF IN ANTIFUNGAL DRUG DISCOVERY

FIELD OF THE INVENTION

The present invention relates to the identification of novel essential fungal specific genes isolated in the yeast pathogen, *Candida albicans* and to their structural and functional relatedness to their *Sacharomyces cerevisiae* counterparts. More specifically the invention relates to the use of these novel essential fungal specific genes in fungal diagnosis and antifungal drug discovery.

BACKGROUND OF THE INVENTION

Opportunistic fungi, including *Candida albicans*, *Aspergillus fumigatus*, *Cryptococcus neoformans*, and *Pneumocystis carinii*, are a rapidly emerging class of microbial pathogens, which cause systemic fungal infection or "mycosis" in patients whose immune system is weakened. Candida spp. rank as the predominant genus of fungal pathogens, accounting for approx. 8% of all bloodstream infections in hospitals today. Alarmingly, the incidence of life-theatening *C. albicans* infections or "candidiasis" have risen sharply over the last two decades, and ironically, the single greatest contributing factor to the prevalence of mycosis in hospitals today is modern medicine itself. Standard medical practices such as organ transplantation, chemotherapy and radiation therapy, suppress the immune system and make patients highly susceptible to fungal infection. Modern diseases, most notoriously, AIDS, also contribute to this growing occurrence of fungal infection In fact, *Pneumocystis carinii* infection is the number one cause of mortality for AIDS victims. Treatment of fungal infection is hampered by the lackof safe and effective antifungal drugs. Antimycotic compounds used today; namely polyenes (amphotericin B) and azole-based derivatives (fluconazole), are of limited efficacy due to the nonspecific toxicity of the former and emmerging

resistance to the latter. Resistance to fluconazole has increased dramatically throughout the decade particularly in *Candida* and *Aspergillus* spp.

Clearly, new antimycotic compounds must be developed to combat fungal infection and resistance. Part of the solution depends on the ellucidation of novel antifungal drug targets (i.e. gene products whose functional inactivation results in cell death). The identification of gene products essential to cell viability in a broad spectrum of fungi, and absent in humans, could serve as novel antifungal drug targets to which rational drug screening can be then employed. From this starting point, drug screens can be developed to identify specific antifungal compounds that inactivate essential and fungal-specific genes, which mimick the validated effect of the gene disruption

Of paramount importance to the antifungal drug discovery process is the genome sequencing projects recently completed for the bakers yeast *Saccharomyces cerevisiae* and under way in *C. albicans*. Although *S. cerevisiae* is not itself pathogenic, it is closely related taxonomically to opportunistic pathogens including *C. albicans*. Consequently, many of the genes identified and studied in *S. cerevisiae* facilitate identification and functional analysis of orthologous genes present in the wealth of sequence information provided by the Stanford *C. albicans* genome project (http://candida.stanford.edu). Such genomic sequencing efforts accelerate the isolation of *C. albicans* genes which potentially participate in essential cellular processes and which therefore could serve as novel antifungal drug targets.

However, gene discovery through genome sequence analysis alone does not validate either known or novel genes as drug targets. Ultimately, target validation needs to be achieved through experimental demonstration of the essentiality of the candidate drug target gene directly within the pathogen, since only a limited concordance exists between gene essentiality for a particular ortholog in different organisms. For example, in a literature search of 13 *C. albicans* essential genes validated by gene disruption, 7 genes (i.e. *CaFKS1*, *CaHSP90*, *CaKRE6*, *CaPRS1*, *CaRAD6*, *CaSNF1*, and *CaEFT2*) are not essential in *S. cerevisiae*. Therefore, although the null phenotype of a gene in one organism may, in some instances, hint at the function of the orthologous

gene in pathogenic yeasts, such predictions can prove invalid after experimentation.

There thus remains a need to identify new essential genes in *C. albicans* and validate same as drug targets.

The present invention seeks to meet these and other needs.

The present description refers to a number of documents, the content of which is herein incorporated by reference in their entirety.

SUMMARY OF THE INVENTION

In general, the present invention relates to essential fungal specific genes that seek to overcome the drawbacks of the prior art associated with targets for antifungal therapy and with the drugs aimed at these targets. In addition, the present invention relates to screening assays and agents identified by same which may display significant specificity to fungi, more particularly to pathogenic fungi, and even more particularly to *Candida albicans*.

The invention concerns essential fungal specific genes in *Candida albicans* and their use in antifungal drug discovery.

More specifically, the present invention relates to the identification of genes known to be essential for viability in *S. cerevisiae* and to a direct assessment of whether an identical phenotype is observed in *C. albicans*. Such genes which are herein found to be essential in *C. albicans* serve as validated antifungal drug targets and provide novel reagents in antifungal drug screening programs.

More specifically, the present invention relates to the nucleic acid and amino acid sequences of *CaKRE5*, *CaALR1* and *CaCDC24* of *Candida albicans*. Furthermore, the present invention relates to the identification of *CaKRE5, CaALR1* and *CaCDC24* as essential genes, thereby validating same as targets for antifungal drug discovery and fungal diagnosis.

Until the present invention, it was unknown whether *KRE5*, *ALR1* and *CDC24* were essential in a wide variety of fungi. While these genes had been shown to be essential in one of budding yeast (e.g. *S. cerevisiae*) and fission yeast (e.g. *S. pombe*), the essentiality of these genes had not been

assessed in a dimorphic or a pathogenic fungi (e.g. *C. albicans*). Thus, the present invention teaches that *KRE5, ALR1* and *CDC24* are essential genes in very different fungi, thereby opening the way to use these genes and gene products as targets for antifungal drug devebpment diagnosis, in a wide variety of fungi, including animal-infesting fungi and plant-infesting fungi. Non-limiting examples of such pathogenic fungi include *Candida albicans, Aspergillus fumigatus, Aspergillus flavus, Aspergillus niger, Coccidiodes immitis, Cryptococcus neoformans, Exophiala dermatitidis, Histoplsma capsulatum, Dermtophytes spp., Microsporum spp., Tricophyton spp., Phytophthora infestans, and Puccinia sorghi.* More particularly, the invention relates to the identification of these genes and gene products as validated drug targets in any organism in the kingdom of Fungi (Mycota). Thus, although the instant description mainly focuses on *Candida albicans*, the present invention may also find utility in a wide range of fungi and more particularly in pathogenic fungi.

Prior to the present invention, the essentiality of these genes had not been verified in an imperfect, dimorphic yeast which survices as an obligate associate of human beings as well as other mammals, such as *Candida albicans*. Moreover, prior to the present invention, there was no reasonable prediction that a null mutation in any one of these three genes in *Candida albicans* would be essential, in view of the significant evolutionary divergence between *C. albicans* and *S. pombe* or *S. cerevisiae* and thus, of the significant difference between the biology of these fungi. For example, in view of the complexity of the pathways in which *KRE5. ALR1* and *CDC24* are implicated, it could not be reasonably predicted that a knockout of *CaKRE5, CaALR1* or *CaCDC24* would not be compensated by other factors, upstream or downstream thereof. C. albicans can become an opportunistic pathogen in immunosuppressed individuals. Its morphology switches from a yeast (budding) form to a pseudohyphal and eventually hyphal (filamentous) morphology depending on particular stimuli. It is generally believed that the hyphal form of *C. albicans* is pathogenic/virulent. Switching from the yeast to hyphal form involves a developmental process referred to as the dimorphic transition.

In a further general aspect, the invention relates to screening assays to identify compounds or agents or drugs to target the essential function of *CaKRE5, CaALR1* or *CaCDC24*. Thus, in a related aspect, the present invention relates to the use of constructs harboring sequences encoding *CaKRE5, CaALR1* or *CaCDC24*, fragments thereof or derivatives thereof, or the cells expressing same, to screen for a compound, agent or drug that targets these genes or gene products.

Further, the invention relates to methods and assays to identify agents which target *KRE5, ALR1* or *CDC24* and more particularly *CaKRE5, CaALR1* or *CaCDC24*. In addition, the invention relates to assays and methods to identify agents which target pathways in which these proteins are implicated.

In accordance with the present invention, there is thus provided in one embodiment, an isolated DNA sequence selected from the group consisting of the fungal specific gene *CaKRE5*, the fungal specific gene *CaALR1*, the fungal specific gene *CaCDC24*, parts thereof, oligonucleotide derived therefrom, nucleotide sequence complementary to all of the above or sequences which hybridizes under high strigency conditions to the above.

In accordance with another embodiment of the present invention, there is provided a method of selecting a compound that modulates the activity of the product encoded by one of *CaKRE5*, or *CaALR1* or *CaCDC24* comprising an incubation of a candidate compound with the gene product, and a determination of the activity of this gene product in the presence of the candidate compound, wherein a potential drug is selected when the activity of the gene product in the presence of the candidate compound is measurably different and in the absence thereof.

In accordance with another embodiment of the present invention, there is provided an isolated nucleic acid molecule consisting of 10 to 50 nucleotides which specifically hybridizes to RNA or DNA encoding *CaKRE5*, *CaALR1, CaCDC24*, or parts thereof or derivatives thereof, wherein nucleic acid molecule is or is complementary to a nucleotide sequence consisting of at least

10 consecutive nucleic acids from the nucleic acid sequence of *CaKRE5*, *CaALR1*, or *CaCDC24*, or derivatives thereof.

In accordance with another embodiment of the present invention, there is provided a method of detecting *CaKRE5*, *CaALR1* or *CaCDC24* in a sample comprising a contacting of the sample with a nucleicacid molecule under conditions that able hybridization to occur between this molecule and a nucleic acid encoding *CaKRE5*, *CaALR1* or *CaCDC24* or parts or derivatives thereof; and detecting the presence of this hybridization.

In accordance with yet another embodiment of the present invention, there is provided a purified *CaKRE5* polypeptide, *CaALR1* polypeptide, or *CaCDC24* polypeptide or epitope bearing portion thereof.

In yet an additional embodiment of the present invention, there is provided an antibody having specific binding affinity to *CaKRE5*, *CaALR1*, *CaCDC24* or an epitope-bearing portion thereof.

More specifically, the present invention relates to the identification and disruption of the *Candida albicans* fungal specific genes, *CaKRE5*, *CaALR1*, and *CaCDC24* which reveal structural and functional relatedness to their *S. cerevisiae* counterparts, and to a validation of their utility in fungal diagnosis and antifungal drug discovery.

As alluded to earlier, while essentiality of *KRE5*, *ALR1* or *CDC24* has been shown in budding or fission yeast, these results cannot be translated to the *C. albicans* system for numerous reasons. For example, while US Patent 5,194,600 teaches the essentiality of the *S. cerevisiae KRE5* gene, a number of observations from fungal biology make it far from obvious as to the presence and/or role of this gene in a pathogenic yeast, of course, the teachings of 5,194,600 are even more remote from teaching or suggesting that a *KRE5* homolog in *C. albicans* would be essential or if it would have utility as an antifungal target. Examples of such observations are listed below.

a) A related gene, *GPT1*, in the yeast *S. pombe* is not essential. Moreover, *GPT1* thought to be involved in protein folding, fails to complement the *S.cerevisiae kre5* mutant, and fails to reduce β-(1,6)-glucan polymer levels in this yeast.

b) The β-(1,6)-glucan polymer could be made in a different way in different yeasts.

c) Genes are lost during evolution and it could thus not be determined a priori whether *C. albicans* retained a *KRE5* related gene. Moreover, the *CaKRE5* fails to complement a *S. cerevisiae kre5* mutant, thus no gene could be recovered by such an approach. Similarly, the DNA sequence of the *C. albicans CaKRE5* gene is sufficiently different from that of *S. cerevisiae*, that it cannot be detected by low stringency Southern hybridization with the *S. cerevisiae KRE5* gene as a probe.

For the purpose of the present invention, the following abbreviations and terms are defined below.

DEFINITIONS

The terminology "gene knockout" or "knockout" refers to a disruption of a nucleic acid sequence which significantly reduces and preferably suppresses or destroys the biological activity of the polypeptide encoded thereby. A number of knockouts are exemplified herein by the introduction of a recombinant nucleic acid molecule comprising one of *CaKRE5, CaALR1* or *CaCDC24* sequences that disrupt at least a portion of the genomic DNA sequence encoding same in *C. albicans*. In the latter case, in which a homozygous disruption (in a diploid organism or state thereof) is present, the mutation is also termed a "null" mutation.

The terminology "sequestering agent" refers to an agent which sequesters one of the validated targets of the present invention in such a manner that it reduces or abrogates the biological activity of the validated target. A non-limiting example of such a sequestering agent includes antibodies specific to one of the validated targets according to the present invention.

The term "fragment", as applied herein to a peptide, refers to at least 7 contiguous amino acids, preferably about 14 to 16 contiguous amino acids, and more preferably, more than 40 contiguous amino acids in length. Such peptides can be produced by well-known methods to those skilled in the art, such as, for example, by proteolytic cleavage, genetic engineering or

chemical synthesis. "Fragments" of the nucleic acid molecules according to the present invention refer to such molecules having at least 12 nt, more particularly at least 18 nt, and even more particularly at least 24 nt which have utility as diagnostic probes and/or primers. It will become apparent to the person of ordinary skill that larger fragments of 100 nt, 1000 nt, 2000 nt and more also find utility in accordance with the present invention.

The terminology "modulation of two factors" is meant to refer to a change in the affinity, strength, rate and the like between such two factors. Having identified *CaKRE5*, *CaALR1* and *CaCDC24* as essential genes and gene products in *C. albicans* opens the way to a modulation of the interaction of these gene products with factors involved in their respective pathways in this fungi as well as others.

Nucleotide sequences are presented herein by single strand, in the 5' to 3' direction, from left to right, using the one letter nucleotide symbols as commonly used in the art and in accordance with the recommendations of the IUPAC-IUB Biochemical Nomenclature Commission.

Unless defined otherwise, the scientific and technological terms and nomenclature used herein have the same meaning as commonly understood by a person of ordinary skill to which this invention pertains. Generally, the procedures for cell cultures, infection, molecular biology methods and the like are common methods used in the art. Such standard techniques can be found in reference manuals such as for example Sambrook et al. (1989, Molecular Cloning - A Laboratory Manual. Cold Spring Harbor Laboratories) and Ausubel et al. (1994, Current Protocols in Molecular Biology. Wiley, New York).

The present description refers to a number of routinely used recombinant DNA (rDNA) technology terms. Nevertheless, definitions of selected examples of such rDNA terms are provided for clarity and consistency.

As used herein, "nucleic acid molecule", refers to a polymer of nucleotides. Non-limiting examples thereof include DNA (e.g. genomic DNA, cDNA) and RNA molecules (e.g. mRNA). The nucleic acid molecule can be obtained by cloning techniques or synthesized. DNA can be double-stranded or single-stranded (coding strand or non-coding strand [antisense]).

The term "recombinant DNA" as known in the art refers to a DNA molecule resulting from the joining of DNA segments. This is often eferred to as genetic engineering.

The term "DNA segment", is used herein, to refer to a DNA molecule comprising a linear stretch or sequence of nucleotides. This sequence when read in accordance with the genetic code, can encode a linear stretch or sequence of amino acids which can be referred to as a polypeptide, protein, protein fragment and the like.

The terminology "amplification pair" refers herein to a pair of oligonucleotides (oligos) of the present invention, which are selected to be used together in amplifying a selected nucleic acid sequence by one of a number of types of amplification processes, preferably a polymerase chain reaction. Other types of amplification processes include ligase chain reaction, strand displacement amplification, or nucleic acid sequence-based amplification, as explained in greater detail below. As commonly known in the art, the oligos are designed to bind to a complementary sequence under selected conditions.

The nucleic acid (e.g. DNA or RNA) for practising the present invention may be obtained according to well known methods.

Nucleic acid fragments in accordance with the present invention include epitope-encoding portions of the polypeptides of the invention. Such portions can be identified by the person of ordinary skill using the nucleic acid sequences of the present invention in accordance with well known methods. Such epitopes are useful in raising antibodies that are specific to the polypeptides of the present invention. The invention also provides nucleic acid molecules which comprise polynucleotide sequences capable of hybridizing under stringent conditions to the polynucleotide sequences of the present invention or to portions thereof.

The term hybridizing to a "portion of a polynucleotide sequence" refers to a polynucleotide which hybridizes to at least 12 nt, more preferably at least 18 nt, even more preferably at least 24 nt and especially to about 50 nt of a polynucleotide sequence of the present invention.

The present invention further provides isolated nucleic acid molecules comprising a polynucleotide sequences which is preferably at least 90% identical, more preferably from 96% to 99% identical, and even more preferably, 95%, 96%, 97%, 98%, 99% or 100% identical to the polynucleic acid sequence encoding the validated targets or fragments and/or derivatives thereof according to the present invention. Methods to compare sequences and determine their homology/identity are well known in the art.

Oligonucleotide probes or primers of the present invention may be of any suitable length, depending on the particular assay format and the particular needs and targeted genomes employed. In general, the oligonucleotide probes or primers are at least 12 nucleotides in length, preferably between 15 and 24 nucleotides, and they may be adapted to be especially suited to a chosen nucleic acid amplification system. As commonly known in the art, the oligonucleotide probes and primers can be designed by taking into consideration the melting point of hydrizidation thereof with its targeted sequence (see below and in Sambrook et al., 1989, Molecular Cloning - A Laboratory Manual, 2nd Edition, CSH Laboratories; Ausubel et al., 1989, in Current Protocols in Molecular Biology, John Wiley & Sons Inc., N.Y.).

The term "oligonucleotide" or "DNA" molecule or sequence refers to a molecule comprised of the deoxyribonucleotides adenine (A), guanine (G), thymine (T) and/or cytosine (C), in a double-stranded form, and comprises or includes a "regulatory element" according to the present invention, as the term is defined herein. The term "oligonucleotide" or "DNA" can be found in linear DNA molecules or fragments, viruses, plasmids, vectors, chromosomes or synthetically derived DNA. As used herein, particular double-stranded DNA sequences may be described according to the normal convention of giving only the sequence in the 5' to 3' direction. "Oligonucleotides" or "oligos" define a molecule having two or more nucleotides (ribo or deoxyribonucleotides). The size of the oligo will be dictated by the particular situation and ultimately on the particular use thereof and adapted accordingly by the person of ordinary skill. An oligonucleotide can be synthetised chemically or derived by cloning according to well known methods.

As used herein, a "primer" defines an oligonucleotide which is capable of annealing to a target sequence. thereby creating a double stranded region which can serve as an initiation point for DNA synthesis under suitable conditions.

The terms "homolog" and "homologous" as they relate to nucleic acid sequences (e.g. gene sequences) relate to nucleic acid sequence from different fungi that have significantly related nucleotide sequences, and consequently significantly related encoded gene products, and preferably have a related biological function. Homologous gene sequences or coding sequences have at least 70% sequence identity (as defined by the maximal base match in a computer-generated alignment of two or more nucleic acid sequences) over at least one sequence window of 48 nucleotides, more preferably at least 80 or 85%, still more preferably at least 90%, and most preferably at least 95%. The polypeptide products of homologous genes have at least 35% amino acid sequence identity over at least one sequence window of 18 amino acid residues, more preferably at least 40%, still more preferably at least 50% or 60%, and most preferably at least 70%, 80%, or 90%. Preferably, the homologous gene product is also a functional homolog, meaning that the homolog will functionally complement one or more biological activities of the product being compared. For nucleotide or amino acid sequence comparisons where a homology is defined by a % sequence identity, the percentage is determined using any one of the known programs as very well known in the art. A non-limiting example of such a program is the BLAST program (with default parameters (Altschul et al., 1997, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acid Res. 25:3389-3402). Any of a variety of algorithms known in the art which provide comparable results can also be used, preferably using default parameters. Performance characteristics for three different algorithms in homology searching is described in Salamov et al., 1999, "Combining sensitive database searches with multiple intermediates to detect distant homologues." *Protein Eng.* 12:95-100. Another exemplary program package is the GCG™ package from the University of Wisconsin.

Homologs may also or in addition be characterized by the ability of two complementary nucleic acid strands to hybridize to each other under appropriately stringent conditions. Hybridizations are typically and preferably conducted with probe-length nucleic acid molecules, preferably 20-100 nucleotides in length. Those skilled in the art understand how to estimate and adjust the stringency of hybridization conditions such that sequences having at least a desired level of complementarity will stably hybridize, while those having lower complementarity will not. For examples of hybridization conditions and parameters, see, *e.g.*,. Sambrook et al. (1989) *supra*; and Ausubel et al. (1994) *supra*.

"Nucleic acid hybridization" refers generally to the hybridization of two single-stranded nucleic acid molecules having complementary base sequences, which under appropriate conditions will form a thermodynamically favored double-stranded structure. Examples of hybridization conditions can be found in the two laboratory manuals referred above (Sambrook et al., 1989, *supra* and Ausubel et al., 1989, *supra*) and are commonly known in the art. In the case of a hybridization toa nitrocellulose filter, as for example in the well known Southern blotting procedure, a nitrocellulose filter can be incubated overnight at 65°C with a labelled probe in a solution containing 50% formamide, high salt (5 x SSC or 5 x SSPE), 5 x Denhardt's solution, 1% SDS. and 100 µg/ml denatured carrier DNA (e.g. salmon sperm DNA). The non-specifically binding probe can then be washed off the filter by several washes in 0.2 x SSC/0.1% SDS at a temperature which is selected in view of the desired stringency: room temperature (low stringency), 42°C (moderate stringency) or 65°C (high stringency). The selected temperature is based on the melting temperature (Tm) of the DNA hybrid.Of course, RNA-DNA hybrids can also be formed and detected. In such cases, the conditions of hybridization and washing can be adapted according to well known methods by the person of ordinary skill. Stringent conditions will be preferably used (Sambrook et al.,1989, *supra*).

Probes of the invention can be utilized with naturally occurring sugar-phosphate backbones as well as modified backbones including

phosphorothioates, dithionates, alkyl phosphonates and α-nucleotides and the like. Modified sugar-phosphate backbones are generally taught by Miller, 1988, Ann. Reports Med. Chem. 23:295 and Moran et al., 1987, Nucleic acid molecule. Acids Res., 14:5019. Probes of the invention can be constructed of either ribonucleic acid (RNA) or deoxyribonucleic acid (DNA), and preferably of DNA.

The types of detection methods in which probes can be used include Southern blots (DNA detection), dot or slot blots (DNA, RNA), and Northern blots (RNA detection). Although less preferred, labelled proteins could also be used to detect a particular nucleic acid sequence to which it binds. Other detection methods include kits containing probes on a dipstick setup and the like.

Although the present invention is not specifically dependent on the use of a label for the detection of a particular nucleic acid sequence, such a label is often beneficial, by increasing the sensitivity of the detection. Furthermore, this increase in sensitivity enables automation. Probes can be labelled according to numerous well known methods (Sambrook et al., 1989, supra). Non-limiting examples of labels include ^{3}H, ^{14}C, ^{32}P, and ^{35}S. Non-limiting examples of detectable markers include ligands, fluorophores, chemiluminescent agents, enzymes, and antibodies. Other detectable markers for use with probes, which can enable an increase in sensitivity of the method of the invention, include biotin and radionucleotides. It will become evident to the person of ordinary skill that the choice of a particular label dictates the manner in which it is bound to the probe.

As commonly known, radioactive nucleotides can be incorporated into probes of the invention by several methods. Non-limiting examples thereof include kinasing the 5' ends of the probes using gamma ^{32}P ATP and polynucleotide kinase, using the Klenow fragment of Pol I of *E. coli* in the presence of radioactive dNTP (e.g. uniformly labelled DNA probe using random oligonucleotide primers in low-melt gels), using the SP6/T7 system to transcribe a DNA segment in the presence of one or more radioactive NTP, and the like.

Amplification of a selected, or target, nucleic acid sequence may be carried out by a number of suitable methods. See generally Kwoh et al..

1990, Am. Biotechnol. Lab. 8:14-25. Numerous amplification techniques have been described and can be readily adapted to suit particular needs of a person of ordinary skill. Non-limiting examples of amplification techniques include polymerase chain reaction (PCR), ligase chain reaction (LCR), strand displacement amplification (SDA), transcription-based amplification, the Qβ replicase system and NASBA (Kwoh et al., 1989, Proc. Natl. Acad. Sci. USA 86, 1173-1177; Lizardi et al., 1988, BioTechnology 6:1197-1202; Malek et al., 1994, Methods Mol. Biol., 28:253-260; and Sambrook et al., 1989, *supra*). Preferably, amplification will be carried out using PCR.

Polymerase chain reaction (PCR) is carried out in accordance with known techniques. See, e.g., U.S. Pat. Nos. 4,683,195; 4,683,202; 4,800,159; and 4,965,188 (the disclosures of all three U.S. Patent are incorporated herein by reference). In general, PCR involves, a treatment of a nucleic acid sample (e.g., in the presence of a heat stable DNA polymerase) under hybridizing conditions, with one oligonucleotide primer for each strand of the specific sequence to be detected. An extension product of each primer which is synthesized is complementary to each of the two nucleic acid strands, with the primers sufficiently complementary to each strand of the specific sequence to hybridize therewith. The extension product synthesized from each primer can also serve as a template for further synthesis of extension products using the same primers. Following a sufficient number of rounds of synthesis of extension products, the sample is analysed to assess whether the sequence or sequences to be detected are present. Detection of the amplified sequence may be carried out by visualization following EtBr staining of the DNA following gel electrophores, or using a detectable label in accordance with known techniques, and the like. For a review on PCR techniques (see PCR Protocols, A Guide to Methods and Amplifications, Michael et al. Eds. Acad. Press. 1990).

Ligase chain reaction (LCR) is carried out in accordance with known techniques (Weiss, 1991, Science 254:1292). Adaptation of the protocol to meet the desired needs can be carried out by a person of ordinary skill. Strand displacement amplification (SDA) is also carried out in accordance with known techniques or adaptations thereof to meet the particular needs (Walker et al.,

1992, Proc. Natl. Acad. Sci. USA 89:392-396; and ibid., 1992, Nucleic Acids Res. 20:1691-1696).

As used herein, the term "gene" is well known in the art and relates to a nucleic acid sequence defining a single protein or polypeptide. A "structural gene" defines a DNA sequence which is transcribed into RNA and translated into a protein having a specific amino acid sequence thereby giving rise to a specific polypeptide or protein. It will be readily recognized by the person of ordinary skill, that the nucleic acid sequence of the present invention can be incorporated into anyone of numerous established kit formats which are well known in the art.

A "heterologous" (e.g. a heterologous gene) region of a DNA molecule is a subsegment segment of DNA within a larger segment that is not found in association therewith in nature. The term "heterologous" can be similarly used to define two polypeptidic segments not joined together in nature. Non-limiting examples of heterologous genes include reporter genes such as luciferase, chloramphenicol acetyl transferase, β-galactosidase, and the like which can be juxtaposed or joined to heterologous control regions or to heterologous polypeptides.

The term "vector" is commonly known in the art and defines a plasmid DNA, phage DNA, viral DNA and the like, which can serve as a DNA vehicle into which DNA of the present invention can be cloned. Numerous types of vectors exist and are well known in the art.

The term "expression" defines the process by which a gene is transcribed into mRNA (transcription), the mRNA is then being translated (translation) into one polypeptide (or protein) or more.

The terminology "expression vector" defines a vector or vehicle as described above but designed to enable the expression of an inserted sequence following transformation into a host. The cloned gene (inserted sequence) is usually placed under the control of control element sequences such as promoter sequences. The placing of a cloned gene under such control sequences is often referred to as being operably linked to control elements or sequences.

Operably linked sequences may also include two segments that are transcribed onto the same RNA transcript. Thus, two sequences, such as a promoter and a "reporter sequence" are operably linked if transcription commencing in the promoter will produce an RNA transcript of the reporter sequence. In order to be "operably linked" it is not necessary that two sequences be immediately adjacent to one another.

Expression control sequences will vary depending on whether the vector is designed to express the operably linked gene in a prokaryotic or eukaryotic host or both (shuttle vectors) and can additionally contain transcriptional elements such as enhancer elements, termination sequences, tissue-specificity elements, and/or translational initiation and termination sites.

Prokaryotic expressions are useful for the preparation of large quantities of the protein encoded by the DNA sequence of interest. This protein can be purified according to standard protocols that take advantage of the intrinsic properties thereof, such as size and charge (e.g. SDS gel electrophoresis, gel filtration, centrifugation, ion exchange chromatography...). In addition, the protein of interest can be purified via affinity chromatography using polyclonal or monoclonal antibodies. The purified protein can be used for therapeutic applications.

The DNA construct can be a vector comprising a promoter that is operably linked to an oligonucleotide sequence of the present invention, which is in turn, operably linked to a heterologous gene, such as the gene for the luciferase reporter molecule. "Promoter" refers to a DNA regulatory region capable of binding directly or indirectly to RNA polymerase in a cell and intiating transcription of a downstream (3' direction) coding sequence. For purposes of the present invention, the promoter is bound at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter will be found a transcription initiation site (conveniently defined by mapping with S1 nuclease), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase. Eukaryotic promoters will often, but not always, contain "TATA"

boxes and "CCAT" boxes. Prokaryotic promoters contain -10 and -35 consensus sequences which serve to initiate transcription and the transcript products contain Shine-Dalgarno sequences, which serve as ribosome binding sequences during translation initiation.

As used herein, the designation "functional derivative" denotes, in the context of a functional derivative of a sequence whether an nucleic acid or amino acid sequence, a molecule that retains a biological activity (either function or structural) that is substantially similar to that of the original sequence. This functional derivative or equivalent may be a natural derivative or may be prepared synthetically. Such derivatives include amino acid sequences having substitutions, deletions, or additions of one or more amino acids, provided that the biological activity of the protein is conserved. The same applies to derivatives of nucleic acid sequences which can have substitutions, deletions, or additions of one or more nucleotides, provided that the biological activity of the sequence is generally maintained. When relating to a protein sequence, the substituting amino acid as chemico-physical properties which are similar to that of the substituted amino acid. The similar chemico-physical properties include, similarities in charge, bulkiness, hydrophobicity, hydrophylicity and the like. The term "functional derivatives" is intended to include "fragments", "segments", "variants", "analogs" or "chemical derivatives" of the subject matter of the present invention.

As well-known in the art, a conservative mutation or substitution of an amino acid refers to mutation or substitution which maintains 1) the structure of the backbone of the polypeptide (e.g. a beta sheet or alpha-helical structure); 2) the charge or hydrophobicity of the amino acid; or 3) the bulkiness of the side chain. More specifically, the well-known terminologies "hydrophilic residues" relate to serine or threonine. "Hydrophobic residues" refer to leucine, isoleucine, phenylalanine, valine or alanine. "Positively charged residues" relate to lysine, arginine or hystidine. Negatively charged residues" refer to aspartic acid or glutamic acid. Residues having "bulky side chains" refer to phenylalanine, tryptophan or tyrosine.

Peptides, protein fragments, and the like in accordance with the present invention can be modified in accordance with well-known methods dependently or independently of the sequence thereof. For example, peptides can be derived from the wild-type sequence exemplified herein in the figures using conservative amino acid substitutions at 1, 2, 3 or more positions. The terminology "conservative amino acid substitutions" is well-known in the art which relates to substitution of a particular amino acid by one having a similar characteristic (e.g. aspartic acid for glutamic acid, or isoleucine for leucine). Of course, non-conservative amino acid substitutions can also be carried out, as well as other types of modifications such as deletions or insertions, provided that these modifications modify the peptide, in a suitable way (e.g. without affecting the biological activity of the peptide if this is what is intended by the modification). A list of exemplary conservative amino acid substitutions is given hereinbelow.

CONSERVATIVE AMINO ACID REPLACEMENTS

For Amino Acid	Code	Replace With
Alanine	A	D-Ala, Gly, Aib, β-Ala, Acp, L-Cys, D-Cys
Arginine	R	D-Arg, Lys, D-Lys, homo-Arg, D-homo-Arg, Met, Ile, D-Met, D-Ile, Orn, D-Orn
Asparagine	N	D-Asn, Asp, D-Asp, Glu, D-Glu, Gln, D-Gln
Aspartic Acid	D	D-Asp, D-Asn, Asn, Glu, D-Glu, Gln, D-Gln
Cysteine	C	D-Cys, S-Me-Cys, Met, D-Met, Thr, D-Thr
Glutamine	Q	D-Gln, Asn, D-Asn, Glu, D-Glu, Asp, D-Asp
Glutamic Acid	E	D-Glu, D-Asp, Asp, Asn, D-Asn, Gln, D-Gln
Glycine	G	Ala, D-Ala, Pro, D-Pro, Aib, β-Ala, Acp
Isoleucine	I	D-Ile, Val, D-Val, AdaA, AdaG, Leu, D-Leu, Met, D-Met
Leucine	L	D-Leu, Val, D-Val, AdaA, AdaG, Leu, D-Leu. Met, D-Met
Lysine	K	D-Lys, Arg, D-Arg, homo-Arg, D-homo-Arg, Met, D-Met, Ile, D-Ile, Orn, D-Orn
Methionine	M	D-Met, S-Me-Cys, Ile, D-Ile, Leu, D-Leu, Val, D-Val
Phenylalanine	F	D-Phe, Tyr, D-Thr, L-Dopa, His, D-His, Trp, D-Trp, Trans-3,4, or 5-phenylproline, AdaA, AdaG, cis-3,4, or 5-phenylproline, Bpa, D-Bpa
Proline	P	D-Pro, L-I-thioazolidine-4-carboxylic acid, D-or L-1-oxazolidine-4-carboxylic acid (Kauer, U.S. Pat. No. (4,511,390)
Serine	S	D-Ser, Thr, D-Thr, allo-Thr, Met, D-Met, Met (O), D-Met(O), L-Cys, D-Cys
Threonine	T	D-Thr, Ser, D-Ser, allo-Thr, Met, D-Met, Met(O), D-Met(O), Val, D-Val
Tyrosine	Y	D-Tyr, Phe, D-Phe, L-Dopa, His, D-His
Valine	V	D-Val, Leu, D-Leu, Ile, D-Ile, Met, D-Met, AdaA, AdaG

As can be seen in this table, some of these modifications can be used to render the peptide more resistant to proteolysis. Of course, modifications of the peptides can also be effected without affecting the primary sequence thereof using enzymatic or chemical treatment as well-known in the art.

Thus, the term "variant" refers herein to a protein or nucleic acid molecule which is substantially similar in structure and biological activity to the protein or nucleic acid of the present invention. Of course, conservedamino acids can be targeted and replaced (or deleted) with a "non-conservative" amino acid in order to reduce, or destroy the biological activity of the protein. Non-limiting examples of such genetically modified proteins include dominant negative mutants.

As used herein, "chemical derivatives" is meant to cover additional chemical moieties not normally part of the subject matter of the invention. Such moieties could affect the physico-chemical characteristic of the derivative (e.g. solubility, absorption, half life and the like, decrease of toxicity). Such moieties are exemplified in Remington's Pharmaceutical Sciences (e.g. 1980). Methods of coupling these chemical-physical moieties to a polypeptide are well known in the art. It will be understood that chemical modifications and the like could also be used to produce inactive or less active agents or compounds. These agents or compounds could be used as negative controls or for eliciting an immunological response. Thus, eliciting immunological tolerance using an inactive modification of one of the validated targets in accordance with the present invention is also within the scope of the present invention.

The term "allele" defines an alternative form of a gene which occupies a given locus on a chromosome.

It should be understood that numerous types of antifungal polypeptides, fragments, and derivatives thereof can be produced using numerous types of modifications of the amino acid chain. Such numerous types of modifications are well-known to those skilled in the art. Broadly, these modifications include, without being limited thereto. a reduction ofthe size of the molecule. and/or the modification of the amino acid sequence thereof. Also,

chemical modifications such as, for example, the incorporation of modified or non-natural amino acids or non-amino acid moieties, can be made to polypeptide derivative or fragment thereof, in accordance with the present invention. Thus, synthetic peptides including natural, synthesized or modified amino acids, or mixtures thereof, are within the scope of the present invention.

Numerous types of modifications or derivatizations of the antifungals of the present invention, and particularly of the validated targets of the present invention, are taught in Genaro, 1995, Remington's Pharmaceutical Science. The method for coupling different moieties to a molecule in accordance with the present invention are well-known in the art. A non-limiting example thereof includes a covalent modification of the proteins, fragments, or derivatives thereof. More specifically, modifications of the amino acids in accordance with the present invention include, for example, modification of the cysteinyl residues of the histidyl residues, lysinyl and aminoterminal residues, arginyl residues, thyrosyl residues, carboxyl side groups, glutaminyl and aspariginyl residues. Other modifications of amino acids can also be found in Creighton, 1983, In Proteins, Freeman and Co. Ed., 79-86.

As commonly known, a "mutation" is a detectable change in the genetic material which can be transmitted to adaughter cell. As well known, a mutation can be, for example, a detectable change in one or more deoxyribonucleotide. For example, nucleotides can be added, deleted, substituted for, inverted, or transposed to a new position. Spontaneous mutations and experimentally induced mutations exist. The result of a mutations of nucleic acid molecule is a mutant nucleic acid molecule. A mutant polypeptide can be encoded from this mutant nucleic acid molecule.

The terminology "dominant negative mutation" refers to a mutation which can somehow sequester a binding partner, such that thebinding partner is no longer available to perform, regulate or affect an essential function in the cell. Hence, this sequestration affects the essential function of the binding partner and enables an assayable change in the cell growth of the cell. In one preferred embodiment, the change is a decrease in growth of the cell, or even death thereof.

As used herein, the term "purified" refers to a molecule having been separated from a cellular component. Thus, for example, a "purified protein" has been purified to a level not found in nature. A "substantially pure" molecule is a molecule that is lacking in most other cellular components.

As used herein, the terms "molecule", "compound" or "ligand" are used interchangeably and broadly to refer to natural, synthetic or semi-synthetic molecules or compounds. The term "molecule" therefore denotes for example chemicals, macromolecules, cell or tissue extracts (from plants or animals) and the like. Non limiting examples of molecules include nucleic acid molecules, peptides, antibodies, carbohydrates and pharmaceutical agents. The agents can be selected and screened by a variety of means including random screening, rational selection and by rational design using for example protein or ligand modelling methods such as computer modelling, combinatorial library screening and the like. It shall be understood that under certain embodiments, more than one "agents" or "molecules" can be tested simultaneously. Indeed, pools of molecules can be tested. Upon the identification of a pool of molecules as having an effect on an interaction according to the present invention, the molecules can be tested in smaller pools or tested individually to identify the molecule initially responsible for the effect. The terms "rationally selected" or "rationally designed" are meant to define compounds which have been chosen based on the configuration of the validated targets or interaction domains thereof of the present invention. As will be understood by the person of ordinary skill, macromolecules having non-naturally occurring modifications are also within the scope of the term "molecule". For example, peptidomimetics, well known in the pharmaceutical industry and generally referred to as peptide analogs can be generated by modelling as mentioned above. Similarly, in a preferred embodiment, the polypeptides of the present invention are modified to enhance their stability. The molecules identified in accordance with the teachings of the present invention have a therapeutic value in diseases or conditions associated with a fungal infection, and particularly with *C. albicans* infections. Alternatively, the molecules identified in accordance with the teachings of the present invention find utility in the development of more efficient antifungal agents.

The term "mimetic" refers to a compound which is structurally and functionally related to a reference compound, whether natural, synthetic or chimeric. The term "peptidomimetic" is a non-peptide or polypeptide compound which mimics the activity-related aspects of the 3-dimensional structure of a peptide or polypeptide. Thus, peptidomimetic can mimic the structure of a fragment or portion of a fungi polypeptide. In accordance with one embodiment of the present invention, the peptide backbone of a mutant of a validated target of the present invention is transformed into a carbon-based hydrophobic structure which retains its antifungal activity. This peptidomimetic compound therefore corresponds to the structure of the active portion of the mutant from which it was designed. Such type of derivatization can be done using standard medical chemistry methods.

Libraries of compounds (publicly available or commercially available) are well-known in the art. The term "compounds" is also meant to cover ribozymes (see, for example, US 5,712,384, US 5,879,938; and 4,987,071), and aptamers (see, for example, US 5,756,291 and US 5,792613).

It will be apparent to a skilled artisan that the present invention is amenable to the chip technology for screening compounds or diagnosing fungi infection. Furthermore, screening assays in accordance with the present invention can be carried out using the well-known array or micro-array technology.

The present invention also provides antisense nucleic acid molecules which can be used for example to decrease or abrogate the expression of the nucleic acid sequences or proteins of the present invention. An antisense nucleic acid molecule according to the present invention refers to a molecule capable of forming a stable duplex or triplex with a portion of its targeted nucleic acid sequence (DNA or RNA). In one particular embodiment, the antisense is specific to 4E-BP1. The use of antisense nucleic acid molecules and the design and modification of such molecules is well known in the art as described for example in WO 96/32966, WO 96/11266, WO 94/15646, WO 93/08845 and USP 5.593,974. Antisense nucleic acid molecules according to the present invention can be derived from the nucleic acid sequences and

modified in accordance to well known methods. For example, some antisense molecules can be designed to be more resistant to degradation to increase their affinity to their targeted sequence, to affect their transport to chosen cell types or cell compartments, and/or to enhance their lipid solubility bu using nucleotide analogs and/or substituting chosen chemical fragments thereof, as commonly known in the art.

It shall be understood that the *"in vivo"* experimental model can also be used to carry out an *"in vitro"* assay. For example, extracts from the indicator cells of the present invention can be prepared and used in one of the *in vitro* method of the present invention or an *in vitro* method known in the art.

As used herein the recitation "indicator cells" refers to cells that express, in one particular embodiment, one of *CaKRE5*, *CaALR1*, and *CaCDC24*, in such a way that an identifiable or selectable phenotype or characteristic is observable or detectable. Such indicator cells can be used in the screening assays of the present invention. In certain embodiments, the indicator cells have been engineered so as to express a chosen derivative, fragment, homolog, or mutant of these interacting domains Preferably, the cells are fungal cells. In one embodiment, the cells are *S. cerevisiae* cells, in another *C. albicans* cells. In one particular embodiment, the indicator cell is a yeast cell harboring vectors enabling the use of the two hybrid system technology, as well known in the art (Ausubel et al., 1994, *supra*) and can be used to test a compound or a library thereof. In one embodiment, a reporter gene encoding a selectable marker or an assayable protein can be operably linked to a control element such that expression of the selectable marker or assayable protein is dependent on a function of one of the validated targets. Such an indicator cell could be used to rapidly screen at high-throughput a vast array of test molecules In a particular embodiment, the reporter gene is luciferase or β-Gal.

In one embodiment, the validated targets of the present invention may be provided as a fusion protein. The design of constructs therefor and the expression and production of fusion proteins are well known in the art (Sambrook et al., 1989, *supra*; and Ausubel et al., 1994, *supra*). In a particular embodiment, both interaction domains are part of fusion proteins. A non-limiting

example of such fusion proteins includes a LexA-X fusion (DNA-binding domain-4E-X; bait, wherein X is a validated target of the present invention or part or derivative thereof) and a B42 fusion (transactivator domain-Y; prey, wherein Y is a factor or part thereof which binds to X). In yet another particular embodiment, the LexA-X and B42-Y fusion proteins are expressed in a yeast cell also harboring a reporter gene operably linked to a LexA operator and/or LexA responsive element. Of course, it will be recognized that other fusion proteins can be used in such 2 hybrid systems. Furthermore, it will be recognized that the fusion proteins need not contain the full-length validated target or mutant thereof or polypeptide with which it interacts. Indeed, fragments of these polypeptides, provided that they comprise the interacting domains, can be used in accordance with the present invention.

Non-limiting examples of such fusion proteins include a hemaglutinin fusions, Gluthione-S-transferase (GST) fusions and Maltose binding protein (MBP) fusions. In certain embodiments, it might be beneficial to introduce a protease cleavage site between the two polypeptide sequences which have been fused. Such protease cleavage sites between two heterologously fused polypeptides are well known in the art.

In certain embodiments, it might also be beneficial to fuse the interaction domains of the present invention to signal peptide sequences enabling a secretion of the fusion protein from the host cell. Signal peptides from diverse organisms are well known in the art. Bacterial OmpA and yeast Suc2 are two non limiting examples of proteins containing signal sequences. In certain embodiments, it might also be beneficial to introduce a linker (commonly known) between the interaction domain and the heterologous polypeptide portion. Such fusion protein find utility in the assays of the present invention as well as for purification purposes, detection purposes and the like.

For certainty, the sequences and polypeptides useful to practice the invention include without being limited thereto mutants, homologs, subtypes, alleles and the like. It shall be understood that in certain embodiments, the sequences of the present invention encode a functional (albeit defective) interaction domain. It will be clear to the person of ordinary skill that

whether an interaction domain of the present invention, variant, derivative, or fragment thereof retains its function in binding to its partner can be readily determined by using the teachings and assays of the present invention and the general teachings of the art.

Of course, the interaction domains of the present invention can be modified, for example by *in vitro* mutagenesis, to dissect the structure-function relationship thereof and permit a better design and identification of modulating compounds. Derivative or analogs having lost their biological function of interacting with their respective interaction may find an additional utility (in addition to a function as a dominant negative, for example) in raising antibodies. Such analogs or derivatives could be used for example to raise antibodies to the interaction domains of the present invention. These antibodies could be used for detection or purification purposes. In addition, these antibodies could also act as competitive or non-competitive inhibitor and be found to be modulators of the activity of the targets of the present invention.

A host cell or indicator cell has been "transfected" by exogenous or heterologous DNA (e.g. a DNA construct) when such DNA has been introduced inside the cell. The transfecting DNA may or may not be integrated (covalently linked) into chromosomal DNA making up the genome of the cell. In prokaryotes, yeast, and mammalian cells for example,the transfecting DNA may be maintained on a episomal element suchas a plasmid. Transfection and transformation methods are well known in the art (Sambrook et al., 1989, *supra*; Ausubel et al., 1994 *supra*; Yeast Genetic Course, A Laboratory Manual, CSH Press 1987).

In general, techniques for preparing antibodies (including monoclonal antibodies and hybridomas) and for detecting antigens using antibodies are well known in the art (Campbell, 1984, In "Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology", Elsevier Science Publisher, Amsterdam, The Netherlands) and in Harlow et al., 1988 (in: Antibody- A Laboratory Manual, CSH Laboratories). The present invention also provides polyclonal, monoclonal antibodies, or humanized

versions thereof, chimeric antibodies and the like which inhibit or neutralize their respective interaction domains and/or are specific thereto.

From the specification and appended claims, the term therapeutic agent should be taken in a broad sense so as to also include a combination of at least two such therapeutic agents.

In one particular embodiment, the present invention provides the means to treat fungal infection comprising an administration of an effective amount of an antifungal agent of the present invention.

For administration to humans, the prescribing medical professional will ultimately determine the appropriate form and dosage for a given patient, and this can be expected to vary according to the chosen therapeutic regimen (e.g. DNA construct, protein, molecule), the response and condition of the patient as well as the severity of the disease.

Composition within the scope of the present invention shoud contain the active agent (e.g. protein, nucleic acid, or molecule) in an amount effective to achieve the desired therapeutic effect while avoiding adverse side effects. Typically, the nucleic acids in accordance with the present invention can be administered to mammals (e.g. humans) in doses ranging from 0.005 to 1 mg per kg of body weight per day of the mammal which is treated. Pharmaceutically acceptable preparations and salts of the active agent are within thescope of the present invention and are well known in the art (Remington's Pharmaceutical Science, 16th Ed., Mack Ed.). For the administration of polypeptides, antagonists, agonists and the like, the amount administered should be chosen so as to avoid adverse side effects. The dosage will be adapted by the clinician in accordance with conventional factors such as the extent of the disease and different parameters from the patient. Typically, 0.001 to 50 mg/kg/day will be administered to the mammal.

BRIEF DESCRIPTION OF THE DRAWINGS

Having thus generally described the invention, reference will now be made to the accompanying drawings, showing by way of illustration a preferred embodiment thereof, and in which:

Figure 1 shows *CaKRE5* sequence and comparison to the *S. cerevisiae KRE5*, *Drosophila melanogaster UGGT1*, and *S. pombe GPT1* encoded proteins. (A) illustrates nucleotide and predicted amino acid sequence of CaKre5p. The CaKre5p signal peptide is underlined in bold. The ER retention sequence His-Asp-Glu-Leu (HDEL) is indicated in bold at the C-terminus. Non-canonical CTG codons encoding Ser in place of Leu are italicized. (B) shows protein sequence alignment between CaKre5p, Kre5p, Gpt1p, and Uggtp. Proteins are shown in single-letter amino acid code with amino acid identities shaded in black and similarities shaded in gray. Gaps introduced to improve alignment are indicated by dashes and amino acid positions are shown at the left;

Figure 2 shows *CaALR1* sequence and comparison to *S. cerevisiae* Alr1p and Alr2p. (A) illustrates nucleotide and predicted amino acid sequence of *CaALR1*. Two hydrophobic amino acid stretches predicted to serve as transmembrane domains are indicated in bold. Non-canonical CTG codons are italicized. (B) shows protein sequence alignment between CaAlr1p, Alr1p, and Alr2p. Proteins are shown in single-letter amino acid code with amino acid identities shaded in black and similarities shaded in gray. Dashes indicate gaps introduced to improve alignment;

Figure 3 shows *CaCDC24* sequence and comparison to *CDC24* from *S. cerevisiae* and *S. pombe*. (A) illustrates nucleotide and predicted amino acid sequence of *CaCDC24*. Non-canonical CTG codons are italicized. (B) shows protein sequence alignment between CaCdc24p, *S. cerevisiae* Cdc24p, and the *S. pombe* homolog, Scd1p. The CaCdc24p dbl homology domain extends from amino acids 280-500. A pleckstrin homology domain is detected from residues 500-700. Protein alignments are formated as described in Fig. 1 and 2; and

Figure 4 illustrates disruption of *CaKRE5*, *CaALR1*, and *CaCDC24*. Restriction maps of (A) *CaKRE5*, (C) *CaALR1*, and (E) *CaCDC24* display restriction sites pertinent to disruption strategies. The insertion position of the *hisG-URA3-hisG* disruption module relative the *CaKRE5*, *CaALR1*, and *CaCDC24* open reading frames (indicated by open arrows) is indicated as well

as probes used to verify disruptions by Southern blot analysis. (B, D, F.) show southern blot verification of targeted integration of the *hisG-URA3-hisG* disruption module into *CaKRE5*, *CaALR1*, and *CaCDC24* and its precise excision after 5-FOA treatment. (B) shows genomic DNA extracted from *Candida albicans* wild-type strain, CAI-4 (lane 1), heterozygote *CaKRE5/cakre5Δ::hisG-URA3-hisG* (lane 2), heterozygote *CaKRE5/cakre5Δ::hisG* after 5-FOA treatment (lane 3), and a representive transformant resulting from the second round of transformation into a *CaKRE5/cakre5Δ::hisG* heterozygote (lane 4), were digested with HindIII and analyzed using *CaKRE5*, *hisG*, and *CaURA3* probes. Asterisks identify the 1.6 kb ladder fragment that nonspecifically hybridizes to the three probes. (D) shows genomic DNA extracted from CAI-4 (lane 1), heterozygote *CaALR1/caalr1Δ::hisG-URA3-hisG* (lane 2), heterozygote *CaALR1/caalr1Δ::hisG* after 5-FOA treatment (lane 3), and a representive transformant resulting from the second round of transformation into a *CaALR1/caalr1Δ::hisG* heterozygote (lane 4), were digested with EcoRI and analyzed using *CaALR1*, *hisG*, and *CaURA3* probes. (F) shows genomic DNA extracted from CAI-4 (lane 1), heterozygote *CaCDC24/cacdc24Δ::hisG-URA3-hisG* containing the disruption module in orientation 1 (lane 2), heterozygote *CaCDC24/cacdc24Δ::hisG-URA3-hisG* containing the disruption module in orientation 2 (lane 3), heterozygote *CaCDC24/cacdc24Δ*::hisG (orientation 1) after 5-FOA treatment (lane 4), heterozygote *CaCDC24/cacdc24Δ::hisG* (orientation 2) after 5-FOA treatment (lane 5) and a representive transformant resulting from the second round of transformation into a *CaALR1/caalr1Δ::hisG* (orientation 1) heterozygote (lane 6), were digested with EcoRI and analyzed using *CaCDC24*, *hisG*, and *CaURA3* probes.

Other objects, advantages and features of the present invention will become more apparent upon reading of the following non-restrictive description of preferred embodiments with reference to the accompanying drawing which is exemplary and should not be interpreted as limiting the scope of the present invention.

DESCRIPTION OF THE PREFERRED EMBODIMENT

Three *C. albicans* genes whose gene products are homologous to those encoded by the essential genes *KRE5, CDC24*, and *ALR1* from *S. cerevisiae* were identified. These genes participate in essential cellular functions of cell wall biosythesis, polarized growth, and divalent cationtransport, respectively. Disruption of these genes in *C. albicans* experimentally demonstrates their essential role in this pathogenic yeast. Database searches fail to identify clear homologous counterparts in *Caenorhabditis elegans*, mouse and *H. sapiens* genomes, supporting the utility of these genes as novel antifungal targets.

Full length clones of *CaKRE5, CaCDC24* and *CaALR1* using available fragments of *C. albicans* DNA were isolated by Polymerase Chain Reaction (PCR) to amplify genomic DNA derived from *C. albicans* strain SC5314. The PCR products were radiolabeled and used to probe the *C. albicans* genomic library by colony hybridization. DNA sequencing revealed complete open reading frames of *CaKRE5, CaCDC24* and *CaALR1* sharing statistically significant homology to their *S. cerevisiae* counterparts namely *KRE5, CDC24* and *ALR1* all of which have met several criteria expected for potential antifungal drug targets.

Disruption of *CaKRE5, CaCDC24* and *CaALR1* was performed. The disruption plasmids were digested and transformed into *C. albicans* strain CA14. Southern blot analysis confirmed that the afoementioned genes are essential in *C. albicans*.

CaKRE5, CaCDC24 and *CaALR1* were used in antifungal screening assays which confirmed their potential to screen for novel antifungal compounds.

KRE5

The *C. albicans KRE5* gene meets several criteria expected for a potential antifungal drug target. In *S. cerevisiae*, deletion of *KRE5* confers a lethal phenotype (2). Although *KRE5*-deleted cells are known to be viable in one particular strain background, they are extremely slow growing and

spontaneous extragenic suppressors are required to propagate *kre5null* cells under laboratory conditions. Genetic analyses suggest that *KRE5*, together with a number of additional *KRE* genes (e.g. *KRE9*) participate in the *in vivo* synthesis of β-(1,6)-glucan. β-(1,6)-glucan covalently cross-links or "glues" other cell surface constituents, namely β-(1,3)-glucan, mannan, and chitin into the final wall structure and has been shown to be essential for viability in both *S. cerevisiae* and *C. albicans* (1,2 and references therein). Importantly, β-(1,6)-glucan has been demonstrated to exist in a number of additional fungal classes including other yeast and filamentous *Ascomycetes*, *Basidiomycetes*, *Zygomycetes* and *Oomycetes*, emphasizing the likelihood that gene products functioning in the β-(1,6)-glucan biosynthetic pathway could serve as broad spectrum drug targets. Moreover, experimental efforts have failed to detect β-(1,6)-glucan in higher eukaryotes, suggesting that inhibitory compounds identified to act against CaKre5p would likely display a minimal toxicity to mammalian and more particularly to humans. Having now shown that *CaKRE5* is essential *C. albicans*, and knowing that *KRE5* is also essential in *S. cerevisiae*, two yeasts which have significantly diverged evolutionarily, strongly suggest that *KRE5* is a target for antifungal drug screening and diagnosis in a wide variety of fungi, including animal- and plant-infesting fungi.

Consistent with a role in β-(1,6)-glucan biosynthesis, *in vivo* levels of this polymer are reduced substantially in *kre5-1* cells versus an isogenic wild type strain, and are completely absent in several independently-suppressed *kre5* null strains (2). In addition, *kre5* mutants show a number of genetic interactions with *KRE6*, another gene involved in β-(1,6)-glucan assembly. Although the biochemistry of β-(1,6)-glucan synthesis remains poorly understood, recent studies demonstrate that cell wall mannoproteins are extensively glucosylated through β-(1,6) linkages and that this modification plays a central role in their anchorage within the extracellular matrix. Kre5p plays a critical role in this process as Cwp1p, an abundent cell wall protein which is demonstrated to be highly glucosylated through β-(1,6)-glucan addition, is undetected in the cell wall fraction of *kre5null* cells, and instead secreted into the medium.

The predicted *KRE5* gene product offers only limited insight into a possible biochemical activity related to β-(1,6)-glucan production. *KRE5* encodes a large secretory protein containing both an N-terminal signal peptide and C-terminal HDEL retention signal for localization to the endoplasmic reticulum. Interestingly, Kre5p has limited but significant homology to UDP-glucose:glycoprotein glycosyltransferases (UGGT), an enzyme class participating in the "quality control" of protein folding. Such UGGT enzymes function to "tag" misfolded ER proteins by reglucosylation of *N*-linked GlcNAc2Man9 core oligosaccharide structures present on misfolded proteins. Proteins labelled in this way are substrates for theER chaperonin, calnexin, which facilitates refolding of the misfolded protein. However, genetic analyses to address the relative involvement of *Kre5p* in glucosylation-dependent protein folding and β-(1,6)-glucan biosynthesis demonstrate that the essential function of *Kre5p* is unrelated to protein folding, and instead relates to its role in β-(1,6)-glucan polymer biosynthesis (3). Although it remains to be demonstrated biochemically, *Kre5p* homology to glycosyltransferases likely reflects its role in the early biosynthesis of this polymer.

ALR1

The product of the *C. albicans* gene, *CaALR1*, also meets several criteria characteristic of a suitable antifungal drug target. In *S. cerevisiae*, *ALR1* is essential for cell viability, although this essentiality is suppressed under growth conditions containing non-physiologically-relevant levels of supplementary Mg^{+2}. *ALR1* encodes a 922 amino acid protein containing a highly charged N-terminal domain and two hydrophobic C-terminal regions predicted to serve as membrane spanning domains anchoring the protein at the plasma membrane. Although such a localization remains to be directly demonstrated, deposition to the cell surface makes Alr1p an attractive drug target in terms of both bioavailability and resistance issues. Alr1p shares substantial homology to two additional *S. cerevisiae* proteins, Alr2p (70% identity) and Ykl064p (34% identity). Both Alr1p and Alr2p share limited similarity to CorA, a Salmonella typhimurium/periplasmic

membrane protein involved in divalent cation transport. Mammalian homologues to *ALR1* have not been detected despite extensive homology searches in metazoan databases (data not shown).

Although *ALR1* was identified in a screen for genes that confer increased tolerance to Al^{+3} when overexpressed, biochemical analyses support a role for *ALR1* in the uptake system for Mg^{+2} and possibily other divalant cations. Mg^{+2} is an essential requirement for bacterial and yeast growth. Uptake of radiolabelled Co^{+2}, an analog of Mg^{+2} for uptake assays, correlates with *ALR1* activity.

CDC24

A third potential antifungal drug target is the product of the *C. albicans* gene, *CaCDC24*. *CDC24* is essential for viability in both *S. cerevisiae* and *S. pombe* (5). *CDC24* has been biochemically demonstrated to encode GDP-GTP nucleotide exchange factor (GEF) activity towards Cdc42p, a Rac/Rho-type GTPase involved in polarization of the actin cytoskeleton. Conditional alleles of *CDC24* shifted to the nonpermisive temperature lack a polarized distibution of actin, and consequentially form large, spherical, unbudded cells in which the normal polarized deposition of cell wall material is disrupted. Eventually, *cdc24* mutants lyse at the restrictive temperature. *CDC24*-dependent activation of *CDC42*, is also required for the activation of the pheromone response signal transduction pathway during mating, and likely participates in the activation of this pathway under conditions that promote pseudohyphal development, since a downstream effector of *CDC42*. *STE20*, is required for hyphal formation. Thus *CDC24* regulates cell wall assembly and the yeast-hyphal dimorphic transition: both key cellular processes and targets being actively pursued in antifungal drug screens.

Cdc24p localizes to the cell cortex concentrating at sites of polarized growth and interacts physically with a number of proteins including Cdc42p, Bem1p, and the heterotrimeric G protein β and γ subunits encoded by *STE4* and *STE18* respectively. Cdc24p shares 24% overall identity to its

S. pombe counterpart, Scd1p. Similar homology has not been found in mammalian database protein searches, although Cdc24p does possess limited homology to a domain of the human exchange protein, dbl, and contains a pleckstrin homology domain, common to several mammalian protein classes. In contrast to Cdc24p, which has limited homology outside of fungi, Cdc42p shares 80-85% identity to mammalian proteins. The fungal-specific character of *CDC24* may be due to its role in hallmark fungal processes like bud formation, pseudohyphal growth, and projection formation during mating, whereas *CDC42* performs highly conserved functions (namely actin polymerization and signal transduction) common to all eukaryotes.

Isolation of *CaKRE5, CaCDC24*, and *CaALR1*.

To isolate full length clones of *CaKRE5*, *CaCDC24*, and *CaALR1*, oligonucleotides were designed according to publicly available fragments of *C. albicans* DNA sequence. Polymerase chain reaction (PCR) using oligonucleotide pairs CAKRE5.1/CAKRE5.2, CaCDC24.1/CaCDC24.2, and CaALR1.1/CaALR1.2 to amplify genomic DNA derived from *C. albicans* strain SC5314 yielded 574, 299, and 379 bp products, respectively. These PCR products were ^{32}P-radiolabeled and used to probe a YEp352-based *C. albicans* genomic library by colony hybridization.

Sequence Information

DNA sequencing of two independent isolates representing putative *CaKRE5* and *CaALR1* clones revealed complete open reading frames (orf) sharing statistically significant homology to their *S. cerevisiae* counterparts (Figs. 1, 2). DNA sequencing of multiple isolates of *CaCDC24* revealed an orf containing strong identity to *CDC24*, but predicted to be truncated at its 3' end. The 3' end of *CaCDC24* was isolated by PCR amplification using one oligonucleotide designed from its most 3' sequence ard a second oligonucleotide which anneals to the YEp352 polylinker allowing amplification of *CaCDC24* C-terminal encoding fragments from this *C. albicans* genomic library. Subcloning and DNA sequencing of a 1.0 kb PCR product

completes the *CaCDC24* open reading frame and reveals its gene product to share strong homology to both Cdc24p and Scd1p (Fig. 3).

CaKRE5

Sequence analysis reveals *CaKRE5* and *KRE5* are predicted to encode similarly-sized proteins (1447 vs 1365 amino acids; 166 vs 156 kDA) sharing significant homology throughout their predicted protein sequences (22% identity, 42% similarity; see Fig. 1). Moreover, like *KRE5*, *CaKRE5* is predicted to possess an amino-terminal signal peptide required for translocation into the secretory pathway, and a C-terminal HDEL sequence which facilitates the retention of soluble secretory proteins within the endoplasmic reticulum (ER). Although CaKre5p is more homologous to *S. pombe* and metazoan UGGT proteins throughout its C-terminal UGGT homology domain than to Kre5p, CaKre5p and Kre5p, are moe related to each other over their remaining sequence (approx. 1100 amino acids). This unique homology between the two proteins as well as a similar null phenotypes (see below) suggest that *CaKRE5* likely serves as the *KRE5* counterpart in *C. albicans*.

CaALR1

CaALR1 encodes a 922 amino acid residue protein sharing strong identity to both *ALR1* (1.0e-180) and *ALR2* (1.0e-179: see Fig.2). Like these proteins. *CaALR1* possesses a C-terminal hydrophobic region which likely functions as two transmembrane anchoring domains. *CaALR1* shares only limited homology. however, to two highly homologous regions common to *ALR1* and *ALR2*; neither the N-terminal 250 amino acids of *CaALR1* nor its last 50 amino acids C-terminal the hydrophobic domain share strong similarity to *ALR1* or *ALR2*. In addition. *CaALR1* possesses two unique sequence extentions within the CorA homology region (one 38 amino acids in length, the other, 16 amino acids long) not found in either *ALR1* or *ALR2*. Protein database searches identify a *S.pombe* hypothetical protein sharing strong homology to

CaALR1 (2.7e-107), however no similarity to higher eukaryotic proteins were detected.

CaCDC24

Sequence analysis of the *CaCDC24* gene product reveals extensive homology to both Cdc24p (1e-93) and Scd1p from *S. cerevisiae* and *S. pombe* respectively (2e-61; see Fig.3) throughout their entire open reading frames. Although limited similarity exists between CaCdc24p (and both Cdc24p and Scd1p) and a large number of metazoan proteins (upto 5e-18), in each case this homology is restricted to the nucleotide exchange domain predicted to span amino acid residues 250-500. Extensive analysis of metazoan databases failed to identify significant homology to either the N-terminal (amino acids 1-250) and C-terminal (amino acids 500-844) regions of CaCdc24p suggesting the *CDC24* gene family is conserved exclusively within the fungal kingdom.

Disruption of *CaKRE5*, *CaALR1*, and *CaCDC24*

Experimental strategy

Disruption of *CaKRE5* was performed using the *hisG-CaURA3-hisG* "*URA*-blaster" cassette constructed by Fonzi and Irwin and standard molecular biology techniques (1, and references within). A *cakre5::hisG-CaURA3-hisG* disruption plasmid was constructed by deleting a 780bp BamH1-BglII DNA fragment from the library plasmid isolate, p*CaKRE5*. and replacing it with a 4.0 kb BamHI-BglII DNA fragment containing the *hisG-CaURA3-hisG* module from pCUB-6. This *CaKRE5* disruption plasmid is deleted of DNA sequence encoding amino acids 971-1231, which encompasses approx. 50% of the UGGT homology domain. This *CaKRE5* disruption plasmid was then digested with SphI prior to transformation.

A *CaALR1* disruption allele was constructed by first subcloning a 7.0 kp *CaALR1* BamHI-SalI fragment from YEp352-library isolate p*CaALR1* into PBSKII+. A 841 bp *CaALR1* HindIII-BglII fragment was then replaced with a 4.0 kb *hisG-CaURA3-hisG* DNA fragment digested with HindII

and BamHI from PBSK-*hisG-CaURA3-hisG* . This *CaALR1* disruption allele, which is lacking DNA sequences encoding amino acids 20-299, was digested using BamHI and SalI prior to transformation.

A *CaCDC24* insertion allele was constructed by first deleting a 0.9 kb KpnI fragment from YEp352-library isolate p*CaCDC24* to remove *CaCDC24* upstream sequence containing BamHI and BglII restriction sites which obstruct the insertion of the *hisG-CaURA3-hisG* module. The 4.0 kb BamHI-BglII *hisG-CaURA3-hisG* fragment from pCUB-6 was then ligated into a unique BglII site. The resulting plasmid possessing an insertion allele within *CaCDC24* at amino acid position 306, was digested with KpnI and SalI prior to transformation.

CaKRE5, CaALR1, and *CaCDC24* disruption plasmids were digested as described above, and transformed into *C. albicans* strain CAI-4 using the lithium acetate method. Transformants were selected as Ura+ prototrophs on YNB + Casa plates. Heterozygous disruptants were identified by PCR (data not shown), verified by Southern blot (see below), and prepared for a second round of gene disruption by selecting for 5-FOA resistance. To assess the null phenotype of each gene, a second round of transformations using heterozygous *CaKRE5/cakre5*, *CaALR1/caalr1*, and *CaCDC24/cacdc24 ura3-* strains were performed as outlined above.

Correct integration of the *hisG-CaURA3-hisG* module into *CaKRE5*, *CaALR1*, and *CaCDC24* and *CaURA3* excision from heterozygous strains was verified by Southern blot analysis using the following probes:

(1a) a 1.25 kb XbaI-Kpn1 fragment digested from p*CaKRE5* containing N-terminal coding sequence of *CaKRE5*;

(1b) a 1.7 kb PCR product containing coding sequence from amino acid 404 and 3' flanking sequences of *CaALR1*;

(1c) a 778 bp PCR product containing *CaCDC24* coding sequence from amino acids 154-430;

(2) a 783 bp PCR product which contains the entire *CaURA3* coding region;

(3) a 898bp PCR product encompassing the entire *Salmonella typhimurium hisG* gene. Genomic DNA from *CaKRE5*-disrupted strains were digested with HindIII and EcoR1 was used to digest genomic DNA from *CaALR1* and *CaCDC24*-disrupted strains.

Results

Southern blot analysis revealed that the *cakre5::hisG-CaURA3-hisG* disruption fragment integrated precisely into the wild type locus (Fig. 4B) after the first round of transformations. Both a 5.0 kb wild type band and a 9.0 kb band diagnostic of the *CaKRE5*-disrupted allele were detected using the *CaKRE5* probe (Fig. 4B). The 9.0 kb band was also detected with both the *hisG* and *CaURA3* probes, confirming disruption of the first *CaKRE5* copy. Successful excision of the *CaURA3* gene by growth on 5-FOA was validated by 1) a predicted shift in size of the *CaKRE5* disruption fragment from 9.0 kb to 6.0 kb when probed with either *CaKRE5* or *hisG* probes; and 2) the inability of the *CaURA3* probe to recognize this fragment and the resulting strain having reverted to ura3- prototrophy.

To determine whether *CaKRE5* is essential, the transformation was repeated in two independently-derived *CaKRE5/cakre5::hisG, ura3-/ura3-* heterozygote strains. A total of 36 Ura+ colonies (24 small and 12 large colonies after 3 days of growth) were analyzed by PCR using oligonucleotides which amplify a 2.5 kb wild-type fragment that spans the BamHI and BglII sites bordering the disrupted r egion. All colonies were shown to contain this 2.5 kb wild-type fragment but to lack the 2.8 kb *cakre5::hisG* allele, consistent with the *cakre5::hisG-CaURA3-hisG* module integrating at the disrupted locus. Southern blot analysis using the 3 different probes independently confirmed 4 such Ura+ transformants as *bonafide CaKRE5/cakre5::hisG-CaURA3-hisG* heterozygotes. If disruption of both copies of the gene was not essential, then 50% of the recovered disruptants would be expected to integrate into the *CaKRE5* locus, giving 50% homologous and 50% heterozygous disruptants. This is the case, for example, when disrupting the second wild-type allele of *CaKRE1*. Indeed, *CaKRE1* was shown not to be

essential in *C. albicans* by this disruption method, since an equal number of heterozygous and homozygous strains resulted from this second round of transformations (data not shown). However, the absence of any homozygous *CaKRE5* disrupted transformants being detected among the 36 Ura+ transformants analyzed in this experiment demonstrates that *CaKRE5* is an essential *C. albicans* gene. It further validates *CaKRE5* and its gene product as a therapeutic target for drug discovery in this pathogen.

CaALR1

Southern blot analysis of *CaALR1* first round transformants confirmed correct integration of the *caalr1::hisG-CaURA3-hisG* disruption module as judged by an appropriately sized disruption band of 5.7 kb, and a wild-type fragment predicted to be >9.0 kb detected by the *CaALR1* probe (Fig. 4D). This 5.7 kb band was also detected with both the *hisG* and *CaURA3* probes, confirming disruption of one copy of *CaALR1*. Southern blotting confirmed excision of the *CaURA3* gene by growth on 5-FOA as the *CaALR1* probe detected an expected 5.0 kb fragment due to the absense of *CaURA3*. Moreover, this 5 kb *caalr::hisG* band was also detected using the *hisG* probe but not with the *CaURA3* probe (Fig. 4D).

Determination of the *CaALR1* null phenotype was performed as described for *CaKRE5*. However, as it has been reported that the inviability of the *ALR1* null mutation in *S. cerevisiae* can be partially suppressed by supplementing the medium with $MgCl_2$. Thus, the second transformation was performed by selecting for Ura+ colonies on 500mM $MgCl_2$-containing medium as well as on standardCasa plates. 35+ colonies of various size (22 of which were isolated from $MgCl_2$ -supplemented plates) were analyzed by PCR to confirm *caalr1::hisG-CaURA3-hisG* integration. The second allele from each of these 35 transformants was determined to be wild-type by PCR using oligonucleotides that span the insertion and produce a wild-type 1.6 kb product as opposed to the larger 1.75 kb product of the *caalr::hisG* allele. Southern blot analysis using the 3 different probes independently confirmed 4 such Ura+ transformants as *CaALR1/caalr1::hisG-CaURA3-hisG* heterozygotes. This

inability to identify any homozygous *CaALR1* disrupted transformant among the 35 Ura+ colonies analyzed. experimentally demonstrates that *CaALR1* is an essential *C. albicans* gene and validates the *CaALR1* gene product as a therapeutic target for drug discovery against this pathogen.

CaCDC24

Southern blot analysis of *CaCDC24* first round transformants using the *CaCDC24* gene probe confirmed the correct integration of the *cacdc24::hisG-CaURA3-hisG* insertion fragment as both 2.55 kb and 3.7 kb fragments, which are diagnostic of the insertional allele, were detected in addition to the 2.2 kb wild-type CaCDC24 fragment (Fig. 4F). Moreover, both 2.55 kb and 3.7 kb fragments were detected using *CaURA3* and *hisG* probes. Excision of *CaURA3* from the resulting heterozygote was verified by: 1) detecting a single 3.3 kb fragment unique to 5-FOA resistant colonies using the *CaCDC24* or *hisG* probes; and 2) the failure to detect this band using the *CaURA3* probe (Fig. 4F).

As previously, a second round of transformations using the above described *CaCDC24* heterozygote was performed. 28+ colonies of various size were analyzed by PCR to confirm *cacdc24::hisG-CaURA3-hisG* integration. The second allele from each of these 28 transformants was determined to be wild-type by PCR using oligonucleotides which span the insertion and produce a wild-type 0.5 kb product rather than the 1.6 kb product of the *caalr::hisG* allele. Southern blot analysis using the 3 different probes independently confirmed 4 such Ura+ transformants as *CaCDC24/cacdc24::hisG-CaURA3-hisG* heterzygotes. The inability to identify a homozygous *CaCDC24* disrupted transformant among these 28 Ura+ colonies analyzed, again demonstrates that *CaCDC24* is an essential *C. albicans* gene and is therefore a third validated drug target suitable for drug discovery against this pathogen

The present invention is illustrated in further detail by the following non-limiting examples.

EXAMPLE 1

In vivo Screening Methods for Specific Antifungal Agents

Having now validated *CaKRE5*, *CaALR1* and *CaCDC24* as drug targets in *Candida albicans*, heterologous expression of *CaKRE5*, *CaALR1*, or *CaCDC24* in *S.cerevisiae kre5*, *alr1* and *cdc24* mutants respectively, allows replacement of the *S. cerevisiae* gene with that of its *C. albicans* counterpart and thus permits screening for specific inhibitors to this *bonafide* drug target in a *S. cerevisiae* background where the additional experimental tractability of the organism permits additional sophistication in screen development. For example, drugs which block CaKre5p in *S. cerevisiae* confer K1 killer toxin resistance, and this phenotype can be used to screen for such compounds. In a particular embodiment, *CaKRE5* can be genetically modified to function in *S. cerevisiae* by replacing its promoter sequence with any strong constitutive *S. cerevisiae* promoters (e.g. *GAL10, ACT1, ADH1*). As *C. albicans* utilizes an altered genetic code, in which the standard leucine-CTG codon is translated as serine, all four codons (or any functional subset thereof) could be modified by site-directed mutagenesis to encode serine residues when expressed in *S. cerevisiae*. Compounds that impair CaKre5p activity in *S. cerevisiae* may be screened using a K1 killer toxin sensitivity assay. Similarly, compounds could be screened which inactivate heterologously-expressed *CaCDC24* and consequently disrupt its association with Rsr1p or Cdc42p in a two hybrid assay. Alternatively, *CaCDC24* function could be monitored in a screen for compounds able to disrupt pseudohyphal formation in a *CaCDC24*-dependent manner. A whole cell drug screening assay based on *CaALR1* function could similary be envisaged. For example, *CaALR1*-dependent influx of $^{57}CO_2$+ in a *S. cerevisiae alr1* mutant suppressed by supplementary Mg^{2+} could be monitored to identify compounds which specifically block the import of divalent cations.

EXAMPLE II

***In vitro* Screening Methods for Specific Antifungal Agents**

1. Use of an *in vitro* assay to synthesize β-(1,6)-glucan.

In such an assay the incorporation of labelled glucose from UDP-glucose into a product that can be immunoprecipitated or immobilized with β-(1,6)-glucan antibodies is measured. The specificity of this synthesis can be established by showing its dependence on CaKre5p, and its digestion with β-(1,6)-glucanase.

Drugs which block this *in vitro* synthesis reaction, block β-(1,6)-glucan synthesis and are candidates for antifungal drugs, some may inhibit Kre5p, others may inhibit other steps in the synthesis of this polymer.

2. Use of a specific *in vitro* assay for CaKre5p.

CaKre5p has amino-acid sequence similarities to UDP-glucose glycoprotein glucosyltransferases (4). The CaKre5p protein can be heterogeneously expressed and/or purified from *Candida albicans* and an *in vitro* assay devised by adding purified GPI-anchored cell wall proteins known to normally contain β-(1,6)-glucan linkages in a *KRE5* wild-type background but absent in *kre5* deleted extracts. Such acceptor substrates could be obtained from available *S. cerevisiae kre5* null extracts suppressed by second site mutations or conditional *kre5* strains (e.g. under control of a regulatable promoter or temperature sensitive mutation). CaKre5p dependent protein glycosylation is measured as radiolabelled incorporation of UDP-glucose into the acceptor substrate purified from the *kre5* null extract. Alternatively, it is possible to screen for compounds that bind to immobilized CaKre5p. For example, scintilation proximity assays (SPA) could be developed in high throughput format to detect compounds which disrupt binding between CaKre5p and radiolabelled UDP-glucose. Alternatively, a SPA-based CaKre5P *in vitro* screen may be employed using a labelled antibody to CaKre5p and screening for compounds able to disrupt the CaKre5p:antiCaKre5p antibody dependent fluorescence Compounds identified in such screens serve as lead compounds in the development of novel antifungal therapeutics.

CDC24 has been biochemically demonstrated to encode a GDP-GTP nucleotide exchange factor (GEF) required to convert Cdc42p to a GTP-bound state. An *in vitro* assay to measure CaCdc24p-dependent activation of Cdc42p could be used to screen for inhibitors of CaCdc24p. This could be accomplished by directly measuring the percentage of GTP versus GDP bound by Cdc42p. Alternatively, Cdc24p function could be determined indirectly by measuring Cdc42p-GTP dependent activation of Ste20p kinase activity.

EXAMPLE III

The use of *CaALR1*, *CaKRE5*, and *CaCDC24* in PCR-based diagnosis of fungal infection

Polymerase chain reaction (PCR) based assays provide a number of advantages over traditional serological testing methodologies in diagnosing fungal infection. Issues of epidemiology, fungal resistance, relability, sensitivity, speed, and strain identfication are limited by the spectrum of primers and probes available. The *CaKRE5*, *CaALR1*, and *CaCDC24* gene sequences enable the design of novel primers of potential clinical use. In addition, as CaAlr1p is thought to localize to the plasma membrane and extend out into the periplasmic space/cell wall, this extracellular domain could act as a serological antigen to which antibodies could be raised and used in serological diagnostic assays.

EXAMPLE IV

Plasmid-based reporter constructs which measure Kre5p, Alr1p, or Cdc24p inactivation

Transcriptional profiling of *kre5*, *alr1*, and *cdc24* mutants in *S. cerevisiae* could identify genes which are transcriptionally induced or repressed specifically under conditions of *KRE5*, *ALR1*, or *CDC24* inactivation or overproduction. The identification of promoter elements from genes responsive to the loss of *KRE5*, *ALR1*, or *CDC24* activity offers practical utility in drug screening assays to identify compounds which specifically

inactivate these targets. For example, a chimeric reporter gene (eg. *lacZ, GFP,*) whose expression would be either induced or repressed by such a promoter would reflect activity of Kre5p, and could be used for high-throughput screening of compound libraries. Further, a group of promoters showing such regulated expression would allow a specific fingerprint or transcriptional profile to be buit for the inhibition or overproduction of the *ALR1, CDC24,* or *KRE5* genes. This would allow a reporter set to be constructed that could be used for high-throughput screening of compound libraries giving a specific tool for screening compounds which inhibit these gene products.

CONCLUSION

The aim of the present invention is to provide the identification and subsequent validation of novel drug targets that can be used in specific enzymatic and cellular assays leading to the discovery of new clinically useful antifungal compounds. Although *KRE5, ALR1* and *CDC24* have previously been identified in the baker's yeast, *S. cerevisiae*, prior to the present invention, it was unknown whether orthologous genes would be identified in the human pathogen *C. albicans*, or whether should they exist, these genes would perform identical or similar functions. The *CaKRE5, CaALR1* and *CaCDC24* genes from *C. albicans* have thus been identified and their utility has been validated as novel antifungal drug targets by experimentally demonstrating their essential nature by gene disruption directly in the pathogen. Although the precise role of these gene products remains to be determined, the current understanding of their cellular functions does enable both *in vitro* and *in vivo* antifungal drug screening assay development. Furthermore, and of importance clinically, genome database searches fail to detect significant homology to these genes in metazoans, suggesting that screening for compounds which inactivate these fungal-specific drug targets are less likely to display toxicity to mammals and particularly to humans. *KRE5* and *CDC24* are unique genes in *S. cerevisiae* and irrespective of their inclusion in gene families in *C. albicans*, they retain an essential function. *ALR1p1* is part of a 3 member gene family in *S. cerevisiae*, and sequence similarity to ALR2p has been identified (Stanford Sequencing

Project), however the essential role of CaALR1p in *C. albicans* and their predicted extracellular location offers the potential to screen for novel antifungal compounds which need not enter the cell, circumventing issues of compound delivery and drug resistance.

Thus, the present invention provides the identification of *CaKRE5*, *CaALR1*, and *CaCDC24* as essential in *Candida albicans* and as fungal-specific validated drug antifungal targets. The present invention also provides the means to use these validated targets to screen for artifungal drugs to Mycota in general and more particularly to a pathogenic yeast such as *Candida albicans*. Thus, the present invention extends in a non-obviousway the use of these genes in a pathogenic fungal species, as targets for screening for drugs specifically directed against fungal pathogens.

Although the present invention has been described hereinabove by way of preferred embodiments thereof, it can be modified, without departing from the spirit and nature of the subject invention as defined in the appended claims.

REFERENCES

1. Lussier et al., 1998, Proc. Natl. Acad. Sci. USA 95:9825-9830.

2. Meaden et al., 1990, Mol. Cell. Biol. 10:3013-3019.

3. Orlean, P., 1997, eds. Pringle, J. R., Broach, J. R., and Jones, E. W. Cold Spring Harbor Lab. Press, Plainview, NY. Vol 3, pp 229-362.

4. Shahinian et al., 1998, Genetics 149:843-856.

5. MacDiarmid et al., 1998, J. Biol. Chem. 273:1727-1732.

6. Pringle et al., 1995, Cold Spring Harbor Symp. Quant. Biol. 60: 729-744.

Claims

WHAT IS CLAIMED IS:

1. An isolated DNA sequence selected from the group consisting of:

a) fungal specific gene of *C. albicans* termed *CaKRE5;*

b) fungal specific gene of *C. albicans termed CaALR1;*

c) fungal specific gene of *C. albicans* termed CaCDC24; '

d) a part or oligonucleotide derived from a), b) or c);

e) a nucleotide sequence complementary to any of the nucleotide sequences of a) - d); and

f) a sequence which hybridizes under high stringency conditions to any of the nucleotide sequences of a) - e).

2. The isolated DNA sequence of claim 1, wherein said sequence of *CaKRE5* is as set forth in Figure 1A.

3. The isolated DNA sequence of claim 1, wherein said sequence of *CaALR1* is as set forth in Figure 2A.

4. The isolated DNA sequence of claim 1, wherein said sequence of *CaCDC24* is as set forth in Figure 3A.

5. A method of selecting a compound that modulates the activity of a protein encoded by said *CaKRE5* of claim 2 comprising:

a) incubating a candidate compound with said protein; and

b) determining the activity of said protein in the presence of said candidate compound,

wherein a potential drug is selected when the activity of said protein in the presence of said candidate compound is measurably different than in the absence thereof.

6. A method of selecting a compound that modulates the activity of a protein encoded by said *CaALR1* of claim 3 comprising:

a) incubating a candidate compound with said protein; and

b) determining the activity of said protein in the presence of said candidate compound,

wherein a potential drug is selected when the activity of said protein in the presence of said candidate compound is measurably different than in the absence thereof.

7. A method of selecting a compound that modulates the activity of a protein encoded by said *CaCDC24* of claim 3 comprising:

a) incubating a candidate compound with said protein; and

b) determining the activity of said protein in the presence of said candidate compound,

wherein a potential drug is selected when the activity of said protein in the presence of said candidate compound is measurably different than in the absence thereof.

8. An isolated nucleic acid molecule consisting of 10 to 50 nucleotides which specifically hybridizes to RNA or DNA of claim 1, 2, 3 or 4, wherein said nucleic acid molecule is or is complementary to a nucleotide sequence consisting of at least 10 consecutive nucleotides from said nucleic acid sequence set forth in Figures 1A, 2A or 3A.

9. A method of detecting *CaKRE5*, *CaALR1* or *CaCDC24* in a sample comprising:

a) contacting said sample with a nucleic acid molecule according to claim 8, under conditions such that hybridization occurs; and

b) detecting the presence of said molecule bound to said *CaKRE5*, *CaALR1* or *CaCDC24* nucleic acid.

10. A purified *CaKRE5* polypeptide or an epitope-bearing portion thereof.

11. A purified *CaALR1* polypeptide or an epitope-bearing portion thereof.

12. A purified *CaCDC24* polypeptide or an epitope-bearing portion thereof.

13. The purified *CaKRE5* polypeptide according to daim 10, comprising an amino acid sequence at least 90% identical to the amino acid sequence as set forth in Figure 1B.

14. The purified *CaALR1* polypeptide according to claim 11, comprising an amino acid sequence at least 90% identical to the amino acid sequence as set forth in Figure 2B.

15. The purified *CaCDC24* polypeptide according to claim 12, comprising an amino acid sequence at least 90% identical to the amino acid sequence as set forth in Figure 3B.

16. An antibody having specific binding affinity to the polypeptide or epitope-bearing portion thereof according to claim 10.

17. A method of screening for a compound having antifungal activity through an interaction with a protein selected from *KRE5*, *ALR1* and *CDC24* comprising:

a) incubating a candidate compound with said protein; and

b) determining one of the activity of said protein or of an assayable or observable property associated with a biological function of said protein in the presence of said candidate compound.

wherein a potential antifungal drug is selected when the activity or assayable or observable property of said protein in the presence of said candidate compound is measurably different than in the absence thereof.

18. The method of claim 17, wherein said antifungal activity is effective against a fungi selected from *Candida albicans, Aspergillus fumigatus, Aspergillus flavus, Aspergillus niger, Coccidiodes immitis, Cryptococcus neoformans, Exophiala dermatitidis, Histoplsma capsulatum, Dermtophytes spp., Microsporum spp., Tricophyton spp., Phytophthora infestans and Puccinia sorghi.*

```
1                                                                                                    GTTTCCAGC
10   AGCAGTACCACCACCACCACGATCAACCGCATATTCAGTTTGGAGCACATTGCCAATTTCAAGATGTTGACAAAGTTGCGTCTATGGCACTACCACCTCTACTGGGAGGATGTGT
125  ACCCTGTATTGATCGTGGAGGATGTCGAGCTCCACAACTGCACGTGGGAGTTTCCGTTCACGCTATCGCAATTCAATTACAACTCCAACATCAGGCGACTTGTGGTGTCGTATGC
240  TGAAGGCAACCCGTTTGCGGTGTCTGAACGGTACAGAGAGTTTTTGCAATATGGAAACGGAGAAGACTTTTCAAGTTTGGAGGAGCTTACGGTCACTGTGGCGAGAGGGAGTCTC
355  AACAGCAGCGTGATGTCACGGTTCATGAACACTGGCAACTTCCCGAGACTAAGAGCATTGCGGGTTCTTGCAAGGGAAGGCGCATACAACCTATCGCATTGGTTTGGAAAGTTGC
470  CGACAAAACAGTACGTTGCGGGTACTAGACATGCAGGTGGATTACGAAGCTCGTGACCGGGAGAGAGCATTGAAGGAGGCCAATAGATACTTTCCATTCCTTGATGTGAAGATAC
585  ATAGACCATAAAAGCACAAGGCTGCGAAATATATACGCGTATAGACTCTACTAATAAACATCCAAACCAGAGTGAAAAAAAAAAATACAACACAAACCAGAAAAAAAACAAACGA
700  ACCACTTACAAGACCCATCTCTACCACAACACCAATGTACTGGGTGCTACTCCTTTTCGTGTCGATATGCATGGCCAACACGGAGACATGCTTGGTACGGGTGCCCGAGTACTAC
815  AATATTGTACCGCACCCGTCACCCATATCCAGGGATGCCAGGTTCAGTCGCGAGCTCCATCGTCTCAACACCACCCACACAGTACTACTAGACTACCCCATTGGATCTATCGACG
930  ACCAGGATATGTCCAACATAATCACAGTCACATACGATACCGTTGCGCAACCACGATCAACACTACTAGTGCGCGTGAACAACTACGGAGACAATACGTTTACGAACGGCGACAT
1045 GCTCAACATTAAGCTATGCTGGCCGGCCACCATGCCGTACGACTTTAGCATTGACCATGTGTATATGCACAGCAACGAGTTGGTTGAGAGTGTGGAGGATGAGTTTGATTTGTAT
1160 GTGGCGGTCACCTACGAGTTCCATGCCTTTAGTTATGACAATGGGAGGTTTTTGCAAGAAGAAACGGCATTGTCCTTCCAATTGTACGTGAACAAATTGCCCAGTAGATTCTTAC
1275 CCATTCCATTGGAGTTGTACGAAACAATCGTGTATTTAGTAGATATCACAATATTCATTGTCTGGAACATCTTGCCATATTTGGTTAAGGGTGTATTAGAAGCCGTGGGGCAGTA
1390 GTGTTGCGTTATATTTTAAGGAAAATAAACAAATGATTTTATCAAGTCGATTAGCCTTATAACATTAGCTGATATGTTTGTTGGTCTATAGGTTTTAATGATATTGTTAGATTTA
1505 AGGTTTGCTTTGTAGCTGGCAAAGTTTAGATGCCAATTCGTTGGGGTCGTGTTCACTACCAATACTGCAGTAAAAACGAGTTTGACTCTTTGTATAATATTTAGCTCATTCGCAG
1620 ATCAAATAATTCGTTCTTTTCTAGGTGCCACACTAGCAAAAGGTTATGGTTAAAGAAGGACACGGTGCATTTCCTGTTCCTAAAGCCAATGACATACCGCCTCCTGCAAATATAA
1735 AGAAGATCGGGGCCTTGAAACGTTTTTCAGATGAAGCTTATGCGAAGTCTTTGCTCTATGATGCAGCAAGATTAGTTGCACCAATAATACACGAGCAGAAGTTCAAAGTTGGAAA
```

FIG. 1A

1850 ATTATATGAAATGTATCCTGATAAGGCGGAACTATGGGGCCTAAATGTCAATCACGGACAAAAGATATACCTAAGGTTAAGAGAACATCACAATGATAAACTGTTTCTCCCCATG
1965 GGTGATATAGTAGGGACCTTACTTCATGAATTAACACACAATTTGTATAGTGCTCACGATAGTAAGTTCTACAAGTTTTTGGACAAACTAAAGTCGAGATACGACGACATACATT
2080 GTAGGGGAGCCAAAACAAAATATTTATGCGAGGAAAACAAGGTTGGTAGAGGTGTATTATTATCCGGAAGTTTAGTATCTGTCAGAGAGCAAAGGCTCAAGGAATTAAGCAAACC
2195 AAAGTTTGCGAATGAAAGCAAAGTTTTAGGACTGAATTCAAAAATTAATAAACCTATCGGTGGCTCGCCAAGGGATCTTAGACAGGCAATTCTAGAGGCGGCAGAGCGTCGGTTG
2310 AGAGATTCAAAATGGTGTCATAGTGAAAATGCAGAAACCGAAAGTGTTCCCAAAGAGGACGAGTACGACACAACTCAGGTGGAGCTTATCGGTCCTACAGAAGGTAAACCAGTTG
2425 GAACATTTGCTAATGATATCATTGATTTAACATCGGACACTGAAGAAACTCCAATTCAACCTGATAACCCGAAACGCCGCATACTCCACGAGATAATTGATTTAACTTCAGATAC
2540 AGAAGACATAGAGCCAACATCACCAGAGGTAATATGTATAGATTAAGTTAAATATAAAGGCAAATATATTGCCAATGTAATACTCTTTTAACAGTGTTGTTCTCGTGCAAGGATT
2655 AAGCACCGAAAAAAAATATGTGGATGCGTTGTTATTAGTTTTACTCTTTGCTTTTTCTGAAAAGAAACATTAACGTGTTCTACTAGTTTGTCACACTACGACACAAGTCCTTGAA

Met Ser Phe Ala Arg Tyr Ile Tyr Tyr Thr Ile Ala Val Ala Val Leu Leu Asn Phe Val Lys Ala Thr Glu Asn Asn Asn Phe Lys 29
2770 ATG TCA TTT GCA AGG TAT ATC TAC TAC ACC ATT GCG GTT GCT GTT TTA TTA AAT TTT GTC AAA GCT ACT GAA AAT AAC AAT TTT AAA
Leu Glu Val Glu Ala Ser Trp Ser Asn Ile Asp Phe Leu Pro Ser Phe Ile Glu Ala Ile Val Gly Phe Asn Asp Ser Leu Tyr Glu 58
2857 CTT GAA GTT GAA GCG TCA TGG AGC AAT ATT GAT TTC CTT CCT AGC TTT ATA GAG GCC ATC GTT GGC TTC AAT GAC TCT TTG TAC GAA
Gln Thr Ile Glu Thr Ile Phe Gly Leu Gly Asp Thr Glu Val Glu Leu Glu Asp Asp Ala Ser Asp Gln Glu Ile Tyr Ser Thr Val 87
2944 CAG ACA ATT GAA ACA ATT TTT GGT TTA GGA GAC ACT GAA GTG GAA TTA GAA GAT GAT GCT TCA GAT CAA GAA ATA TAT TCT ACC GTG
Ile Asn Ser Leu Gly Leu Thr Asp Gln Asp Leu Asp Phe Ile Asn Phe Asp Leu Thr Asn Lys Lys His Thr Pro Arg Ile Ala Ala 116
3031 ATC AAC TCA TTA GGG TTA ACA GAT CAA GAT TTG GAT TTT ATT AAT TTT GAT TTA ACC AAC AAA AAA CAT ACA CCA AGA ATC GCA GCC
His Tyr Asp His Tyr Ser Asp Val Leu Thr Lys Phe Gly Asp Arg Leu Lys Ser Glu Cys Ala Lys Asp Ser Phe Gly Asn Ala Val 145
3118 CAT TAC GAT CAC TAT TCT GAT GTT CTA ACT AAG TTT GGC GAT CGA CTC AAA AGT GAA TGT GCA AAA GAC TCT TTT GGG AAT GCA GTG
Glu Thr Lys Asn Gly Gln Ile Gln Thr Trp Leu Leu Tyr Asn Asp Lys Ile Tyr Cys Ser Ala Asn Asp Leu Phe Ala Leu Arg Thr 174

FIG. 1A (cont.)

```
3205 GAA ACG AAA AAT GGT CAA ATT CAA ACG TGG TTA CTA TAT AAC GAT AAG ATA TAT TGT TCG GCT AAT GAT TTG TTT GCA TTA CGA ACT
     Asp Leu Ser Ser His Ser Thr Leu Leu Phe Asp Arg Ile Ile Gly Lys Ser Lys Asp Ala Pro Leu Val Ile Leu Tyr Gly Ser Pro   203
3292 GAT TTG AGT TCT CAT TCT ACA CTT TTA TTT GAT AGG ATT ATT GGA AAA TCA AAA GAT GCA CCT TTG GTG ATT TTA TAT GGA AGC CCG
     Thr Glu Glu Leu Thr Lys Asp Phe Leu Lys Ile Leu Tyr Pro Asp Ala Lys Ala Gly Lys Leu Lys Phe Val Trp Arg Tyr Ile Pro   232
3379 ACT GAG GAA CTG ACT AAA GAT TTT CTT AAA ATA TTG TAT CCA GAT GCA AAG GCT GGA AAA TTA AAG TTT GTA TGG AGG TAC ATT CCA
     Leu Gly Ile Lys Lys Leu Asp Ser Ile Ser Gly Tyr Gly Val Ser Leu Lys Met Glu Lys Tyr Asp Tyr Ser Gly Ala Glu Gly Asn   261
3466 CTG GGA ATC AAA AAA CTG GAC TCA ATT TCT GGA TAC GGT GTA TCA TTG AAA ATG GAA AAG TAT GAT TAT TCT GGT GCA GAA GGA AAT
     Pro Lys Tyr Asp Leu Ser Arg Asp Phe Thr Arg Ile Asn Asp Ser Gln Glu Leu Val Leu Val Asn Glu Lys His Ser Tyr Glu Leu   290
3553 CCA AAG TAT GAT TTG AGT CGA GAT TTC ACC AGA ATT AAT GAC TCG CAA GAG TTG GTC CTG GTC AAT GAA AAA CAT TCG TAT GAA CTT
     Gly Val Lys Leu Thr Ser Phe Ile Leu Ser Asn Arg Tyr Lys Ser Thr Lys Tyr Asp Leu Leu Asp Thr Ile Leu Thr Asn Phe Pro   319
3640 GGT GTT AAA TTG ACT TCA TTC ATA TTA TCC AAT CGT TAC AAG AGT ACT AAA TAT GAC CTT TTA GAT ACG ATT TTA ACC AAC TTT CCC
     Lys Phe Ile Pro Tyr Ile Ala Arg Leu Pro Lys Leu Leu Asn His Glu Lys Val Lys Ser Lys Val Leu Gly Asn Glu Asp Ile Gly   348
3727 AAG TTT ATT CCT TAC ATT GCA CGA TTA CCA AAA TTA CTA AAT CAT GAA AAA GTT AAA TCC AAA GTG CTT GGA AAT GAA GAT ATA GGG
     Leu Ser Gln Asp Ser Tyr Gly Ile Tyr Ile Asn Gly Ser Pro Ile Asn Pro Leu Glu Leu Asp Ile Tyr Asn Leu Gly Thr Arg Ile   377
3814 CTA TCT CAA GAC TCC TAC GGA ATA TAT ATC AAC GGT TCC CCA ATA AAT CCA CTA GAG TTA GAT ATT TAC AAT CTA GGT ACC AGG ATA
     Lys Glu Glu Leu Gln Thr Val Lys Asp Leu Val Lys Leu Gly Phe Asp Thr Val Gln Ala Lys Leu Leu Ile Ala Lys Phe Ala Leu   406
3901 AAG GAG GAA TTA CAG ACT GTG AAA GAT TTA GTG AAA CTT GGA TTT GAT ACC GTA CAA GCA AAG CTC TTG ATA GCA AAA TTT GCT TTA
     Leu Ser Ala Val Lys Gln Thr Gln Phe Arg Asn Gly Asn Thr Leu Met Gly Asn Asn Glu Asn Arg Phe Lys Val Tyr Glu Asn Glu   435
3988 CTT TCA GCT GTT AAA CAA ACA CAA TTT CGA AAT GGG AAT ACA TTA ATG GGT AAC AAT GAA AAT AGA TTT AAA GTG TAT GAA AAT GAA
```

FIG. 1A (cont.)

Phe Lys Lys Gly Ser Ser Glu Lys Gly Gly Val Leu Phe Phe Asn Asn Ile Glu Leu Asp Asn Thr Phe Lys Glu Tyr Thr Thr Asp 464
4075 TTT AAG AAG GGT AGT TCA GAA AAG GGT GGG GTC TTG TTT TTC AAT AAC ATT GAA TTA GAC AAC ACA TTC AAG GAG TAC ACC ACT GAT
Arg Glu Glu Ala Tyr Leu Gly Val Gly Ser His Lys Leu Lys Pro Asn Gln Ile Pro Leu Leu Lys Glu Asn Ile His Asp Leu Ile 493
4162 CGT GAG GAG GCA TAT TTA GGA GTT GGT TCT CAT AAA CTT AAG CCA AAT CAA ATT CCG TTA TTG AAA GAG AAC ATC CAT GAT TTA ATT
Phe Ala Leu Asn Phe Gly Asn Lys Asn Gln Leu Arg Val Phe Phe Thr Leu Ser Lys Val Ile Leu Asp Ser Gly Ile Pro Gln Gln 522
4249 TTC GCA TTA AAT TTT GGG AAC AAA AAC CAA TTG CGG GTG TTT TTC ACT TTA TCT AAG GTG ATT TTG GAC TCC GGT ATA CCT CAA CAA
Val Gly Val Leu Pro Val Ile Gly Asp Asp Pro Met Asp Leu Leu Leu Ala Glu Lys Phe Tyr Trp Ile Ala Glu Lys Ser Ser Thr 551
4336 GTT GGA GTT TTG CCC GTT ATA GCA GAT GAC CCA ATG GAT CTG TTA CTC GCT GAG AAA TTT TAT TGG ATT GCT GAG AAA TCA AGC ACA
Gln Glu Ala Leu Ala Ile Leu Tyr Lys Tyr Phe Glu Ser Asn Ser Pro Asp Glu Val Asp Asp Leu Leu Asp Lys Val Glu Val Pro 580
4423 CAA GAG GCA TTA GCA ATA TTG TAT AAA TAT TTT GAA TCA AAC AGT CCA GAT GAA GTT GAT GAC TTA TTA GAT AAA GTG GAA GTA CCC
Glu Asp Tyr Lys Val Asp Tyr Asn His Val Leu Asn Lys Phe Ser Ile Ser Thr Ala Ser Val Ile Phe Asn Gly Val Ile Tyr Asp 609
4510 GAA GAT TAT AAA GTG GAT TAT AAT CAT GTG TTA AAC AAG TTT TCT ATA TCA ACT GCT TCG GTC ATT TTC AAT GGG GTT ATT TAC GAT
Leu Arg Ala Pro Asn Trp Gln Ile Ala Met Ser Lys Gln Ile Ser Gln Asp Ile Ser Leu Ile Lys Thr Phe Leu Arg Gln Gly Pro 638
4597 TTA AGA GCA CCA AAC TGG CAG ATT GCA ATG AGT AAA CAA ATA TCC CAG GAC ATT TCA CTT ATT AAA ACT TTC TTG AGA CAG GGA CCA
Ile Glu Gly Arg Leu Lys Asp Val Leu Tyr Ser Asn Ala Lys Ser Glu Arg Asn Leu Arg Ile Ile Pro Leu Glu Pro Ser Asp Ile 667
4684 ATA GAG GGT AGA TTG AAA GAT GTT CTT TAC TCT AAT GCA AAA TCA GAA CGC AAT TTA CGT ATA ATT CCA TTA GAA CCT AGT GAC ATT
Ile Tyr Lys Lys Ile Asp Lys Glu Leu Ile Asn Asn Ser Ile Ala Phe Lys Lys Leu Asp Lys Ala Gln Gly Val Ser Gly Thr Phe 696
4771 ATT TAC AAG AAA ATC GAC AAG GAA TTA ATA AAC AAT TCA ATT GCA TTC AAG AAG CTA GAT AAA GCG CAG GGT GTG TCT GGA ACA TTT
Trp Leu Val Ser Asp Phe Thr Lys Ser Ala Ile Ile Thr Gln Leu Ile Asp Leu Leu Leu Leu Leu Lys Lys Lys Ala Ile Gln Ile 725

FIG. 1A (cont.)

```
4858 TGG CTA GTG TCG GAT TTT ACC AAG TCA GCA ATA ATT ACT CAA TTG ATA GAT TTG TTA TTG CTT CTC AAA AAG AAA GCA ATT CAG ATA
      Arg Ile Ile Asn Thr Gly Asp Thr Asp Val Phe Gly Lys Leu Lys Thr Lys Phe Lys Leu Thr Ala Leu Thr Asn Gly Gln Ile Asp   754
4945 AGA ATT ATT AAT ACT GGG GAT ACA GAT GTT TTT GGA AAA TTG AAA ACA AAG TTT AAA TTA ACC GCC TTA ACA AAT GGA CAA ATT GAT
      Glu Ile Ile Glu Ile Leu Lys Lys Ser Asn Ala Ser Ser Ala Asn Asn Asp Glu Leu Lys Lys Met Leu Glu Thr Lys Gln Leu Pro   783
5032 GAA ATT ATT GAG ATT TTG AAA AAA TCC AAC GCT TCA AGT GCA AAT AAT GAT GAA TTG AAA AAA ATG CTT GAG ACT AAG CAA TTA CCT
      Ala His His Ser Phe Leu Leu Phe Asn Ser Arg Tyr Phe Arg Leu Asp Gly Asn Phe Gly Tyr Glu Glu Leu Asp Gln Ile Ile Glu   812
5119 GCT CAT CAC TCT TTT TTG CTA TTC AAC TCT AGA TAT TTT AGA TTG GAT GGA AAT TTT GGA TAC GAG GAA TTG GAT CAA ATT ATA GAG
      Phe Glu Val Ser Gln Arg Leu Asn Leu Ile Pro Asp Ile Met Glu Ala Tyr Pro Asp Glu Phe Arg Ser Lys Lys Val Ser Asp Phe   841
5206 TTT GAA GTA TCT CAA AGA TTG AAC TTA ATC CCG GAC ATC ATG GAG GCA TAT CCG GAT GAG TTT AGG TCG AAG AAG GTA AGT GAT TTT
      Asn Leu Val Leu Ser Gly Leu Asp Asn Met Asp Trp Phe Asp Leu Val Thr Ser Ile Val Thr Lys Ser Phe His Val Asp Glu Lys   870
5293 AAT CTG GTT TTG TCT GGA TTA GAC AAT ATG GAC TGG TTT GAT TTG GTG ACT TCC ATA GTG ACA AAA TCA TTC CAT GTC GAC GAA AAA
      Arg Phe Ile Val Asp Val Asn Arg Phe Asp Phe Ser Ser Leu Asp Phe Ser Asn Ser Ile Asp Val Thr Thr Tyr Glu Glu Asn Ser   899
5380 AGG TTT ATT GTT GAT GTT AAC AGG TTT GAT TTT AGC TCA TTG GAT TTT TCA AAC TCG ATT GAT GTA ACG ACT TAT GAA GAA AAT AGT
      Pro Val Asp Val Leu Ile Ile Leu Asn Pro Met Asp Glu Tyr Ser Gln Lys Leu Ile Ser Leu Val Asn Ser Ile Thr Asp Phe Leu   928
5467 CCA GTT GAT GTA TTA ATA ATT TTG AAC CCT ATG GAT GAA TAT TCT CAA AAA TTG ATA AGC CTT GTT AAT AGC ATT ACA GAT TTT CTG
      Phe Leu Asn Ile Arg Ile Leu Leu Gln Pro Arg Val Asp Leu Lys Glu Glu Ile Lys Ile His Lys Phe Tyr Arg Gly Val Tyr Pro   957
5554 TTC TTG AAC ATT AGA ATC TTA CTA CAA CCA AGA GTG GAT CTG AAA GAA GAG ATC AAA ATT CAC AAG TTT TAT CGT GGT GTG TAT CCT
      Gln Pro Thr Pro Lys Phe Asp Ser Asn Gly Lys Trp Ile Gln His Tyr Ser Ala Gln Phe Glu Ser Ile Pro Ser Asn Val Thr Tyr   986
5641 CAA CCG ACT CCC AAA TTT GAT TCC AAT GGC AAG TGG ATC CAA CAT TAT TCA GCT CAA TTT GAA AGT ATT CCA TCC AAT GTG ACC TAT
```

FIG. 1A (cont.)

```
     Ser Thr Glu Leu Asp Val Pro His Lys Trp Ile Val Val Pro Gln Leu Ser Ser Met Asp Leu Asn Thr Ile Asn Phe Ser Glu Ser  1015
5728 TCT ACT GAA TTA GAT GTT CCA CAT AAG TGG ATA GTT GTT CCT CAA CTG AGT TCG ATG GAT TTA AAC ACA ATC AAT TTC AGC GAA AGC
     His Ser Val Asp Ala Lys Tyr Ser Leu Lys Asn Ile Leu Ile Glu Gly Tyr Ala Arg Asp Ile His Thr Gly Lys Ala Pro Asp Gly  1044
5815 CAC TCT GTT GAT GCA AAA TAC TCT CTA AAA AAT ATA TTA ATT GAA GGA TAT GCT AGA GAT ATT CAT ACT GGG AAG GCC CCT GAT GGT
     Leu Ile Phe Arg Ala Phe Asn Lys Asn Tyr Ser Thr Asp Thr Leu Val Met Thr Ser Leu Asp Tyr Phe Gln Ile Lys Ala Tyr Pro  1073
5902 TTA ATC TTT AGA GCC TTT AAT AAA AAT TAC TCA ACT GAT ACT TTG GTG ATG ACT TCC TTG GAC TAT TTT CAA ATC AAA GCG TAT CCT
     Ser Ile Phe Asn Phe Ser Thr Thr Ser Asn Asp Thr Leu Leu Ser Ala Ser Glu Asn Lys Tyr Gln Ala Asn Thr Glu Glu Leu Glu  1102
5989 AGT ATT TTC AAC TTT AGT ACG ACC TCA AAT GAC ACA TTA TTG TCT GCA TCG GAA AAC AAA TAT CAG GCT AAT ACC GAG GAA TTG GAG
     Ser Ile Glu Val Pro Val Phe Lys Ile Asp Gly Ser Thr Ile Tyr Pro Arg Val Met Lys Ser Gly Asn Asn Lys Pro Met Leu Thr  1131
6076 AGC ATT GAG GTG CCA GTT TTT AAA ATT GAT GGA TCG ACC ATA TAT CCA AGG GTA ATG AAA TCT GGC AAC AAT AAG CCA ATG CTG ACG
     Arg Lys His Ala Asp Ile Asn Ile Phe Thr Ile Ala Ser Gly Gln Leu Tyr Glu Lys Leu Thr Ser Ile Met Ile Ala Ser Val Arg  1160
6163 AGA AAA CAT GCA GAT ATA AAT ATT TTT ACA ATT GCT AGT GGC CAA CTT TAT GAA AAG TTA ACT AGC ATT ATG ATT GCG TCA GTA AGA
     Lys His Asn Pro Ser Leu Thr Ile Lys Phe Trp Ile Leu Glu Asp Phe Val Thr Pro Gln Phe Lys His Leu Val Glu Leu Ile Ser  1189
6250 AAA CAT AAC CCT AGC CTG ACA ATA AAA TTC TGG ATT TTG GAA GAT TTT GTG ACC CCA CAA TTC AAA CAC TTG GTA GAG CTT ATC TCA
     Ile Lys Tyr Asn Val Glu Tyr Glu Phe Ile Ser Tyr Lys Trp Pro Asn Phe Leu Arg Lys Gln Lys Thr Lys Glu Arg Met Ile Trp  1218
6337 ATA AAG TAT AAT GTC GAA TAT GAG TTT ATT AGT TAC AAA TGG CCC AAT TTC TTG AGA AAA CAG AAA ACC AAA GAA AGA ATG ATT TGG
     Gly Tyr Lys Ile Leu Phe Leu Asp Val Leu Phe Pro Gln Asp Leu Asn Lys Ile Ile Phe Ile Asp Ala Asp Gln Ile Cys Arg Ala  1247
6424 GGG TAT AAG ATT TTG TTT TTG GAC GTT TTG TTC CCA CAA GAT CTC AAC AAG ATT ATA TTC ATT GAC GCC GAT CAA ATA TGT AGG GCA
     Asp Leu Thr Glu Leu Val Asn Met Asp Leu Glu Gly Ala Pro Tyr Gly Phe Thr Pro Met Cys Asp Ser Arg Glu Glu Met Glu Gly  1276
```

FIG. 1A (cont.)

6511 GAT TTG ACA GAA TTG GTT AAC ATG GAT CTT GAA GGT GCT CCA TAT GGA TTT ACT CCA ATG TGT GAT TCT CGG GAA GAA ATG GAA GGT

Phe Arg Phe Trp Lys Glu Gly Tyr Trp Ser Asp Val Leu Lys Asp Asp Leu Lys Tyr His Ile Ser Ala Leu Phe Val Val Asp Leu 1305

6598 TTC AGA TTT TGG AAA GAA GGA TAC TGG TCC GAT GTT TTG AAG GAT GAT TTG AAA TAT CAT ATT AGT GCA TTA TTT GTT GTT GAT TTG

Gln Lys Phe Arg Ser Ile Lys Ala Gly Asp Arg Leu Arg Ala His Tyr Gln Lys Leu Ser Ser Asp Pro Asn Ser Leu Ser Asn Leu 1334

6685 CAA AAG TTC AGA TCT ATA AAA GCT GGA GAC AGA TTG AGA GCA CAC TAT CAA AAG CTT TCT AGT GAT CCA AAT TCG TTG AGC AAT TTA

Asp Gln Asp Leu Pro Asn Asn Met Gln Arg Leu Ile Lys Ile Phe Ser Leu Pro Gln Asn Trp Leu Trp Cys Glu Thr Trp Cys Ser 1363

6772 GAT CAA GAT TTG CCC AAT AAT ATG CAA AGA CTG ATA AAA ATT TTC AGT TTG CCT CAA AAT TGG CTC TGG TGT GAA ACG TGG TGC TCA

Asp Lys Ser Leu Glu Asp Ala Lys Met Ile Asp Leu Cys Asn Asn Pro Leu Thr Arg Glu Asn Lys Leu Asp Ala Ala Lys Arg Leu 1392

6859 GAT AAA AGC TTG GAA GAT GCA AAA ATG ATT GAT CTT TGC AAC AAT CCA TTA ACT AGA GAA AAT AAA TTA GAT GCT GCT AAG AGA TTG

Ile Pro Glu Trp Ile Glu Tyr Glu Gln Glu Ile Glu Pro Leu Val Ser Leu Val Gln Asn Asn Thr Ala Lys Glu Val Val Gln Glu 1421

6946 ATC CCA GAA TGG ATT GAA TAC GAG CAA GAA ATT GAA CCA TTG GTA TCA TTA GTA CAG AAT AAT ACC GCC AAA GAA GTT GTT CAA GAG

Ile Glu Ile Asp Thr Asp Gly Glu Gln Glu Glu Gln Lys Gln Glu Ser Asn Asp Asp Asp Phe Ile **His Asp Glu Leu** Stop 1417

7033 ATA GAA ATT GAT ACA GAC GGA GAA CAA GAA GAA CAA AAA CAA GAA AGT AAT GAT GAT GAT TTT ATT CAC GAT GAA TTG TAA TTGTCAA

7121 AGTCACATGGAATAAATAGTGAGAACTCCTGAAACGGCATTAAATACGCACGTTGGGTAGAGATAATACAAATATAGATAAATAGATAGAGAGAAAAAATGTTGGATTTTTTC

7236 AGACTTCTCTTCCTCCTGGGCGCCTCCGGTTTAACTATAATTTTTTAAGATTACACAAAATTCAAGTACACGCACTTTCTAATTATTTTATTGAAGAGTCATAATCAGTAATGAA

7351 TTTTTTTTTTTTTTGATTTTCGATTCTCGATTTCCGATTTCCTCGTTGATTGGTATAATCTAAACGAACAAACAGGTATAAACCTTTGTAGTTAGTTTTTTTTTTTTTCTTTCTTT

7466 CTTTCTTGTACTTTTTCTTAATTGTCTTTTCTTTTTTTCACTTTTCTTAAACTTGTTATATCATTGCCTTAAGACTATTGAATCAGTTCAGTT

FIG. 1A (cont.)

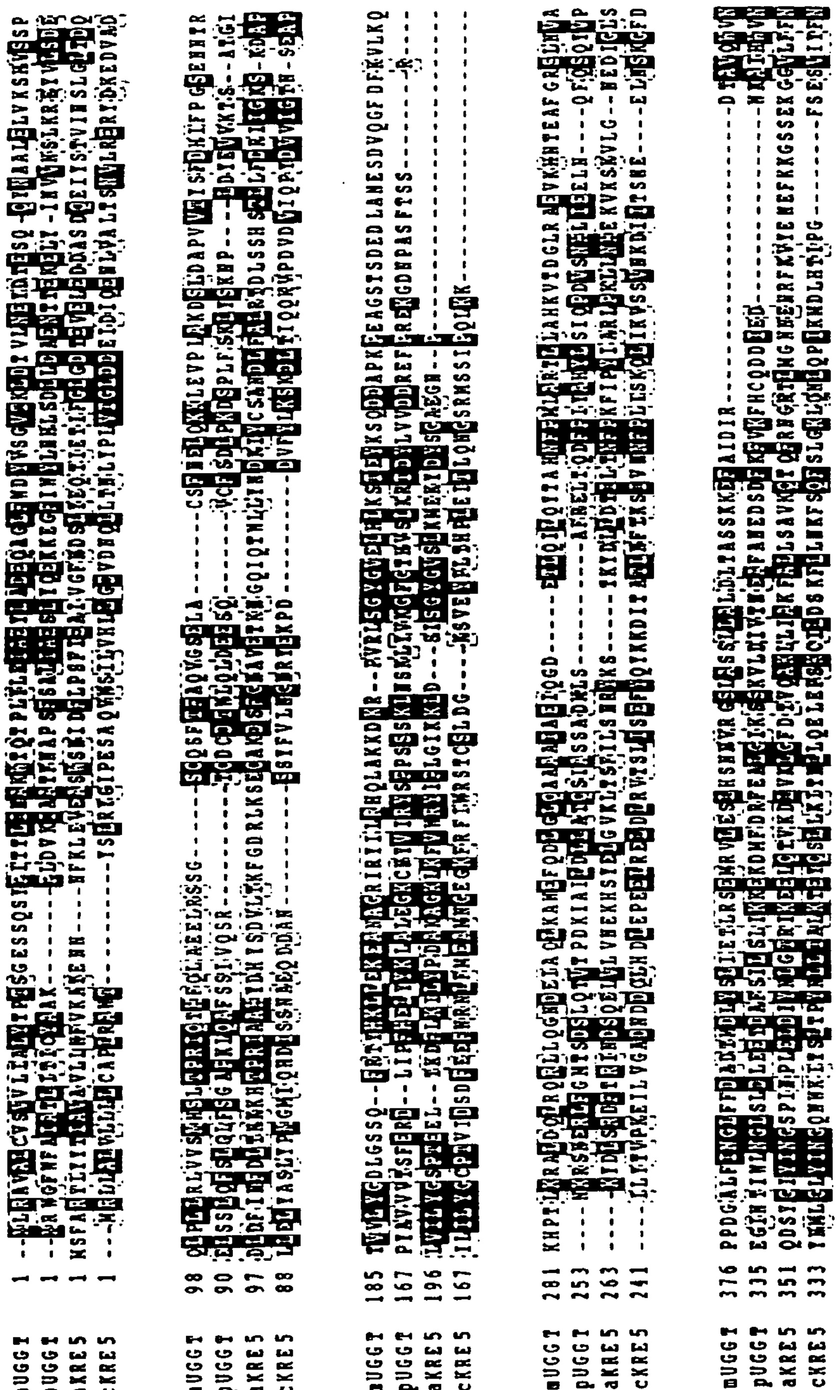

FIG - 1B

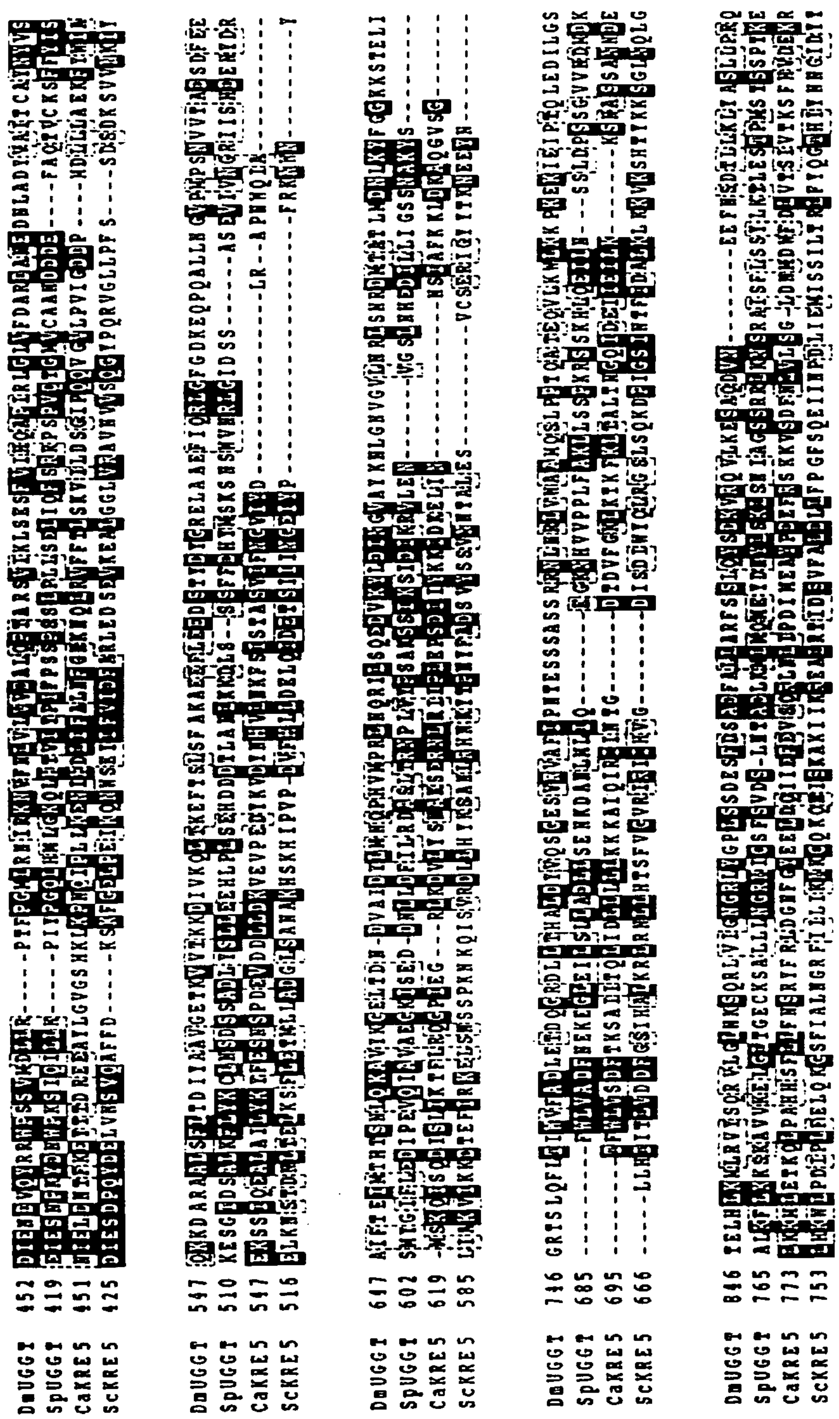

FIG. 1B (cont.)

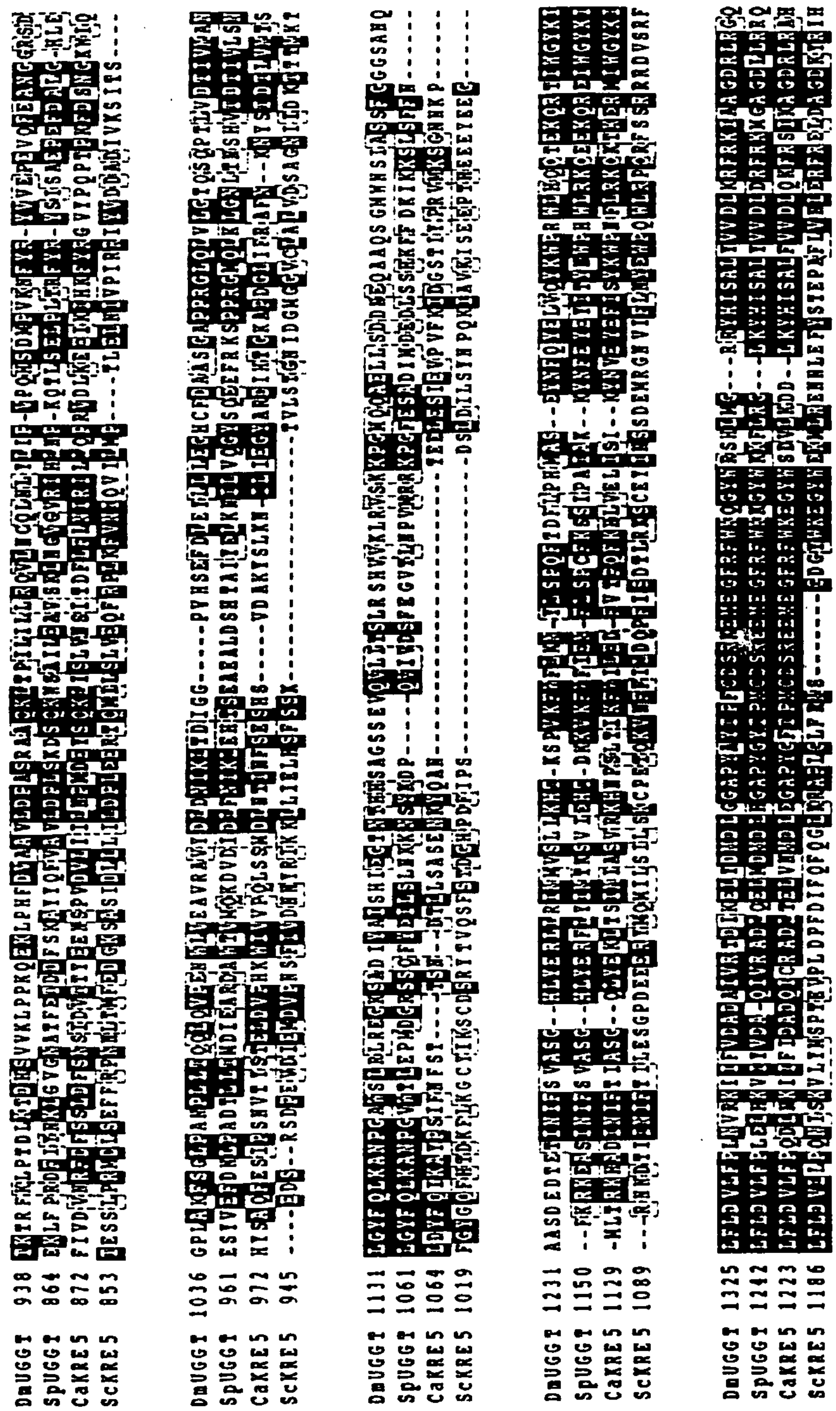

FIG. 1B (cont.)

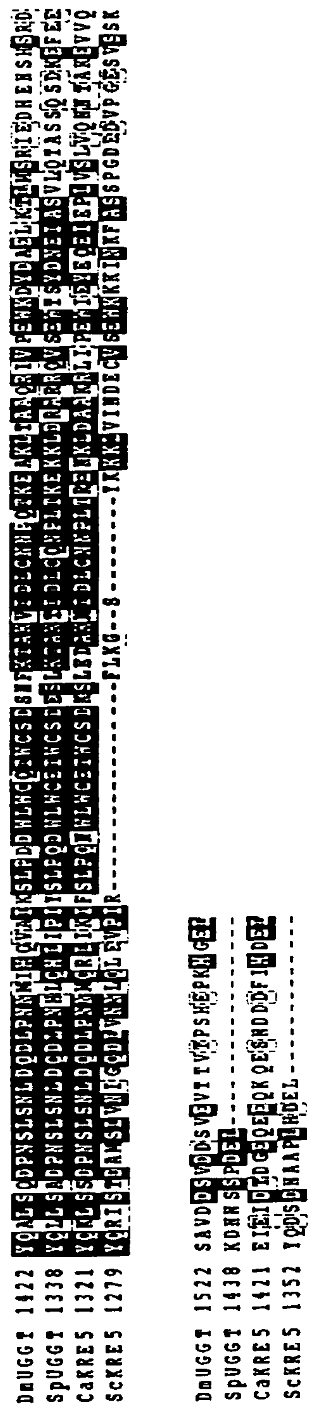
DmUGGT 1422
SpUGGT 1338
CaKRE5 1321
ScKRE5 1279
DmUGGT 1522
SpUGGT 1438
CaKRE5 1421
ScKRE5 1352

FIG. 1B (cont.)

1 TATATAATATATAATATATAAACATAACATATTAAAAAAAACTTCATATTTTCAACTGTTCAACCCAACACCCCCCTGTTCCTAGTCTAGTCCAATTTATTAAGTCA

108 TATATTGTTGATTAATTAAACAGTCACTAGTCCTCAGTCCTCAGTCCTTAGTTCTTGGTTCTTAATATTAAGATNTTCCATTTTTTTTTTTTACCCAAGCTATGAAAATTATTTT

223 TGTTGTCTAACAACTATAATAATATTTACCAGAAATTGCTACAAATATAAATAAATAAATAAATAAATATAATTAAGAGTATATCTCCCCTTTTGTTTTTTTTTTCTTCCCAGCC

Met Ser Asp Ser Glu Ser Tyr Tyr Gln Asn Ser Thr Thr Asn Gln Pro Ile Pro Arg Ser Asp Glu Val Leu Asp Asp His Arg Asn 29

338 ATG TCC GAT AGT GAA AGT TAT TAT CAA AAT TCA ACT ACT AAT CAA CCT ATT CCT AGA TCT GAT GAA GTA TTG GAT GAT CAT AGA AAT

Gln Ile Thr Asn Asp Cys Ala Ile Ser Asp Ser Glu Asp Glu Leu Glu Leu Lys Ser Glu Leu Glu Ser Glu Val Val Lys Ser Glu 58

425 CAA ATC ACT AAT GAT TGT GCC ATT AGT GAT AGT GAA GAT GAG TTG GAA TTA AAA TCA GAA TTA GAA TCA GAA GTT GTA AAA AGC GAA

Lys Gln Gln Gln His His Gln Glu Ile Thr Ser Asp Asn Ala Lys Pro Leu Thr Arg Lys Ser Gly Ser Ser Ile Lys Lys Lys Ser 87

512 AAA CAA CAA CAA CAT CAT CAA GAG ATT ACA TCA GAT AAT GCT AAA CCA TTG ACT CGT AAA TCT GGT TCT TCA ATT AAG AAA AAA TCT

Asn Leu Thr Asp Lys Asp Arg Ile Thr Asn Pro Met Ser Leu Ser Gly Gly Asp Asp Thr Ile Asn Ser Gly His Lys Asn Arg Asn 116

599 AAT CTT ACC GAT AAA GAT AGA ATT ACC AAC CCT ATG AGT TTA TCT GGT GGT GAT GAT ACT ATT AAC AGC GGT CAC AAA AAT CGT AAT

Tyr Asn Met Ser Ser Leu Arg Lys Asp Phe Tyr Leu Lys Asp Asn Thr Asp Asp Asn Ser Thr Asn Asn His Thr His Leu Ala Ile 145

686 TAT AAC ATG AGT TCA TTA CGT AAA GAT TTT TAT TTA AAA GAT AAT ACT GAC GAC AAT TCT ACT AAT AAT CAT ACT CAT CTT GCA ATT

Pro Ile Pro Ile Pro Ile Pro Thr Pro Ile Ile Thr Asn Ala Asn Lys Ser Arg Arg Lys Ser Gln Leu Glu Asn Leu Pro Pro Leu 174

773 CCA ATT CCA ATT CCA ATT CCA ACC CCA ATT ATT ACT AAT GCT AAT AAA TCA AGA AGA AAA TCT CAA TTG GAA AAT TTA CCT CCA TTA

Ile Lys Lys Lys Thr Ile Gly Arg Asn Asn Ser Asn Asn Phe Glu Asn Asp Leu Val Ser Pro Met Thr Lys Met Lys Thr Asn Asp 203

860 ATT AAA AAG AAA ACA ATT GGT CGT AAT AAT TCT AAT AAT TTT GAA AAT GAT TTA GTT AGT CCC ATG ACA AAA ATG AAA ACT AAT GAT

Ser Glu Asp Ile Thr Asn Thr Ser Thr Thr Ala Asn His Met Lys Leu Gly Ile Gly Ala Thr Thr Leu Gly Val Gly Thr Gly Thr 232

947 AGT GAA GAT ATT ACT AAT ACT AGC ACC ACT GCT AAT CAT ATG AAA CTT GGT ATT GGT GCT ACA ACC CTT GGT GTT GGA ACT GGT ACT

Thr Ala Thr Ala Thr Ala Thr Ala Ala Ala Gly Arg Arg Pro Ser Arg Ser Ser Ile Asp Ser Glu Ala Asp Ser His Ala Ser Arg 261

1034 ACC GCC ACT GCC ACT GCC ACT GCT GCT GCT GGT AGA AGA CCA TCT CGT TCA TCT ATT GAT AGT GAA GCT GAT TCT CAT GCA TCA AGA

Ser Ser Gln Glu Thr Glu Glu Asp Val Cys Phe Pro Met Val Gly Asp His Ile Arg Val Asn Gly Ile Asp Phe Asp Glu Ile Asp 290

1121 TCA TCT CAA GAA ACT GAA GAA GAT GTT TGT TTT CCT ATG GTT GGT GAT CAT ATT AGA GTT AAT GGA ATT GAT TTT GAT GAA ATT GAT

FIG. 2A

```
     Glu Phe Ile Arg Glu Glu Arg Glu Glu Ala Tyr Leu Gln Lys Gln Met Ile Ala Lys Asn Ile Leu Arg Ile Asp Glu Phe Gln Asn   319
1208 GAA TTT ATT AGA GAA GAA AGA GAA GAA GCT TAT TTA CAA AAA CAA ATG ATT GCT AAA AAT ATT CTG CGT ATT GAT GAA TTT CAA AAT
     Leu Ser Lys Asn Asn Thr Thr Ser Gly Ala Ser Arg His Pro Tyr His His His Ser Asn Asn Asn Lys Lys Asn Asn Gly Gly Asp   348
1295 CTT TCC AAA AAT AAT ACT ACT AGT GGT GCA TCT CGT CAT CCA TAT CAT CAT CAC AGT AAT AAT AAT AAA AAA AAT AAT GGT GGT GAT
     Gly Gly Gly Ser Ser Met Ala Ala Leu Lys Tyr Thr Pro Lys Asn Ile Leu Lys Lys Thr Leu Ser Arg Phe Glu Phe Thr His Glu   377
1382 GGT GGT GGT TCT AGT ATG GCA GCA TTA AAA TAT ACT CCA AAA AAT ATT TTA AAG AAA ACA TTA TCA AGA TTT GAA TTT ACT CAT GAA
     Asn Ser Ser Ser Ser Glu Glu Ile Tyr Glu Leu Lys Thr Lys Gln Gln Pro Pro Tyr Lys Tyr Asp Asp Gln Leu Ser Leu Thr Ser   406
1469 AAT TCT TCA TCT TCA GAA GAA ATT TAT GAA TTG AAG ACT AAA CAA CAA CCA CCT TAC AAA TAT GAT GAT CAA TTA TCA TTA ACT TCA
     Ser Thr Ser Ser Thr Ser Gly Ser Gly Ser Gly Gln Val Lys Phe Gly Gly Ala Arg Ile Ser Asp Gly Ile Asn Gly Gly Ser Leu   435
1556 TCT ACA TCT TCT ACT TCT GGA TCT GGA TCT GGG CAG GTG AAA TTT GGT GGA GCA AGA ATT TCT GAT GGG ATT AAT GGA GGT TCA TTA
     Pro Asp Arg Phe Ser Leu Phe His Ser Glu Ser Glu Glu Thr Ile His Ala Pro Asp Ile Pro Ser Leu Val Ser Pro Gly Gln Ser   464
1643 CCT GAT AGA TTT TCA CTT TTC CAT TCT GAA TCA GAA GAA ACT ATT CAT GCC CCC GAT ATT CCA TCA TTA GTA TCA CCA GGT CAA TCT
     Val Arg Asp Leu Phe Arg Asn Gly Glu Glu Thr Trp Trp Leu Asp Cys Thr Cys Pro Thr Asp Ser Glu Met Lys Met Leu Ala Lys   493
1730 GTT CGA GAT TTA TTT AGA AAT GGT GAA GAA ACT TGG TGG TTA GAT TGT ACT TGT CCT ACT GAT TCG GAA ATG AAA ATG TTG GCC AAA
     Ala Phe Gly Ile His Pro Leu Thr Ala Glu Asp Ile Arg Met Gln Glu Thr Arg Glu Lys Val Glu Leu Phe Lys Ser Tyr Tyr Phe   522
1817 GCA TTT GGT ATT CAT CCT TTA ACT GCT GAA GAT ATT CGA ATG CAA GAA ACT CGT GAA AAA GTT GAA TTA TTT AAA AGT TAT TAT TTT
     Val Cys Phe His Thr Phe Glu Ala Asp Lys Glu Ser Glu Asp Tyr Leu Glu Pro Ile Asn Val Tyr Ile Val Val Phe His Asp Gly   551
1904 GTT TGT TTC CAT ACT TTT GAA GCT GAT AAA GAA TCT GAA GAT TAT TTA GAA CCG ATA AAT GTT TAT ATT GTT GTT TTC CAT GAT GGT
     Ile Leu Thr Phe His Phe Ser Pro Ile Ser His Pro Ala Asn Val Arg Arg Arg Val Arg Gln Leu Arg Asp Tyr Val Asp Val Ser   580
1991 ATA TTA ACG TTC CAT TTT TCA CCA ATT TCT CAT CCA GCA AAT GTT AGA AGA AGA GTT CGT CAA TTG AGA GAT TAT GTC GAT GTT AGT
     Ala Asp Trp Leu Cys Tyr Ala Leu Ile Asp Glu Ile Thr Asp Gly Phe Ala Pro Val Ile His Gly Ile Glu Tyr Glu Ala Asp Ala   609
2078 GCT GAT TGG TTA TGT TAT GCC TTA ATC GAT GAA ATT ACC GAT GGT TTT GCC CCC GTG ATT CAT GGA ATT GAA TAT GAA GCT GAT GCC
     Ile Glu Asp Ala Val Phe Thr Ala Arg Asp Thr Asp Phe Ser Ser Met Leu Gln Arg Ile Gly Glu Ser Arg Arg Lys Val Met Thr   638
2165 ATT GAA GAT GCC GTT TTC ACT GCT AGA GAT ACT GAT TTT AGT AGT ATG TTA CAA AGA ATT GGT GAA TCA AGA AGA AAA GTC ATG ACT
     Leu Met Arg Leu Leu Ser Gly Lys Ala Asp Val Ile Lys Met Phe Ala Lys Arg Cys Gln Glu Glu Ala Asn Ser Ser Ser Gly Tyr   667
```

FIG. 2A (cont.)

14/24

2252 TTA ATG AGA TTA TTA TCA GGT AAA GCT GAT GTC ATT AAA ATG TTT GCT AAA AGA TGT CAA GAA GAA GCT AAT TCT TCT TCT GGT TAT

Tyr Gln Arg Gln Tyr Asn Leu Gln Gln Gln Gln Gln Gln Ala Pro Pro Pro Pro Pro Asn Pro Ile Ile Thr Ser Pro Ile Asn Ser 696

2339 TAT CAA CGT CAA TAT AAC TTA CAA CAA CAA CAA CAA CAG GCC CCA CCA CCA CCA CCT AAT CCT ATT ATT ACT TCA CCA ATT AAT TCA

Thr Leu Asn Leu Asn Ser Leu Gly Thr Ser Thr Gly Gly Gly Val Gly Val Gly Gly Ile Asn Phe Gly Pro Asn Pro Thr Gly Asn 725

2426 ACT TTG AAT CTT AAT AGT TTA GGA ACT TCA ACT GGT GGA GGA GTA GGA GTA GGA GGA ATT AAT TTT GGT CCC AAT CCA ACT GGA AAT

Asn Thr Asn Thr Asn Thr Asn Thr Thr Gly Ser Pro Ser Pro Pro Gln Gln Gln Gln Gln His Gly Ile Thr Asn Lys Ser Phe Pro 754

2513 AAT ACT AAT ACT AAT ACT AAT ACT ACT GGT TCA CCT TCA CCA CCT CAA CAA CAA CAA CAA CAT GGT ATC ACT AAC AAA TCT TTC CCC

Ile Pro Asp Ala Arg Pro Arg Ala Asp Ile Ala Leu Tyr Leu Gly Asp Ile Gln Asp His Ile Ile Thr Met Phe Gln Asn Leu Leu 783

2600 ATC CCC GAT GCA CGT CCA AGA GCT GAT ATT GCA TTA TAT TTA GGT GAT ATT CAA GAT CAT ATA ATC ACC ATG TTT CAA AAT TTA TTA

Ala Tyr Glu Lys Ile Phe Ser Arg Ser His Ser Asn Tyr Leu Ala Gln Leu Gln Val Glu Ser Phe Asn Ser Asn Asn Lys Ile Thr 812

2687 GCC TAT GAA AAA ATT TTC AGT CGT TCA CAT TCA AAT TAT TTA GCT CAA TTA CAA GTT GAA TCA TTC AAT TCC AAT AAT AAA ATC ACC

Glu Met Phe Ser Lys **Ile Thr Leu Ile Gly Thr Met Leu Val Pro Leu Asn Leu Val Thr Gly Leu Phe Gly** Met Asn Val Arg Val 841

2774 GAA ATG TTT TCT AAA ATT ACT TTG ATT GGG ACA ATG TTA GTT CCA TTA AAT TTA GTC ACG GGA CTT TTT GGT ATG AAT GTA AGA GTC

Pro Gly Glu Gly Gly Thr Asn Leu **Gly Trp Phe Phe Gly Ile Val Gly Val Leu Ile Phe Ile Ile Ile Gly Ser Phe Ile Phe Ala** 870

2861 CCT GGT GAA GGT GGT ACC AAT TTA GGT TGG TTT TTC GGA ATT GTT GGA GTA TTA ATA TTT ATA ATT ATT GGA TCA TTT ATA TTT GCT

Gln Trp Trp Leu Lys Lys Leu Asn Asn Ser Ile Glu Gly Gln Asn Asn Gly Asn Arg Pro Ile Phe Asn His Ser Ser Arg Arg Ser 899

2948 CAA TGG TGG TTG AAA AAA TTG AAT AAT TCA ATT GAA GGA CAA AAT AAT GGT AAT CGA CCA ATT TTT AAT CAT TCA TCA AGA AGA TCA

Ile Arg Ser Leu Gly Leu Lys Lys His Gly Gly Asn Lys Ser Ile Ile Ser Phe Pro Asn Lys Tyr Glu Stop 922

3035 ATT AGA AGT TTA GGT TTA AAA AAA CAT GGT GGT AAT AAA TCA ATT ATT AGT TTC CCC AAT AAA TAT GAA TAA GAATAATCAAAGAAATGCC

3126 ACAGAGTTTGATGGTTTGTTTTTTTTTTTTTATTGTCATGATGGAGTTGTATATACATATACTTTTTATAGAAGTAACAATAGTAAATGATAATAGTAGTCATCAATCATCATA

3241 TTTATAATTGTATATAATCGTATACTAACTTCTTCTTGATTTAGGGAAAGAGTTATATTATTTACTATAAACATTTATTTTTACGAGTTGTGTTAAATTGGAGAGTCAAATTAAT

3356 AGGATGTAAAAGAAGTTTTTAAAGAAGGAATAAAGAAATATTATAATTCAGANGTTCATACAGAAGGGGGGGGAAGGAGAAGGGGATATATATCGGCATTTGTTGGTACTTTTGT

3471 TTTTGAAATAAAATATAAGTTTATCTAAATTATTATCAATTATTATCAATATTGC

FIG. 2A (cont.)

15/24

FIG. 2B

FIG - 2B (cont.)

1 TTATATAACATGATTNGAATAACAATGGTATANTTGAATCCCAAAGAATAT

52 TGAATTTGTATGGGATTGGTAACATAAGATCATTATTGGTGTGGTCAGGGAATGGGACAAGATAAAATTGAAGTGACTCNTGAACTTCAAAAACAAGTTGTTGAGAATATTTTCC

167 AATTGATACCCCATGGCAACCCAGAAGGAATTTCTCAAGCTGAATGGTTGAAATTTTATGAAAATGGTGGCAGATTACCTGATTTGGGATATGGGCCAGGTCACCATTTAGGATT

282 TGAGGAAGAATATGAAGAACATCATTGGAGAAAATACCATGCTAATGATGATCCTGATGTCAAAATAAAACATAAAGAAGATATCGAACATGAGTTGTTACATCATCAACAAGAG

397 ATTGAAGAGACTCACGATAGATCGAGTCGATTGAAAAAATTTGCTGGTGATTACTGGTCAGAGATCAATATCGACAATCTTAAACCAAAATATAGAAACAACAGAAATAGTGAG

512 AGGTTAGTAAAGAGAAACCTTAAAAACAACCAGGAAGCCAACNCTTAGCTTGGCTATATAGAAGTAGAAATAGAAATCAAGATCAATAAAAAGACGCACCAACAGCGCTACTACT

627 ACTACTACATTTTGAAAACAATTTGCTGGGAAGTTGAATGATTTGCACATGTGAGCCACATAGCTTAGGTATAGGTTGCTTGAACTAAACACCGAAGTTCTTGTATTAAAGATAG

742 AATTCCTCTCTCTCTCAAAAAGCTGTCAAATTGACACACCAGTAGAAGTATTTTTGACGGGTATATTGGCAAGTCAAATTGAATGTGTGTCATTTCAATCCTACAAAAAGATACA

857 AAAATGAATAAACGAATTAATATAATAGTTGTCATCATCGTCAAATAATAGCAGAAAATTACATTATCCTATGGATTGTAATGTGATTATGATAATGGTTGTAATTGAGGTCGTA

972 GTAGTAGTAGTCGTGTGTTTCATTTTATGCAACATTAACCACAATAACAAGAACAATAGAGGGGGGGAGGGAATTGATTGTCTTACTAGTTTTCTATTACAAAGAATTTCTTTTG

1087 TTTTAGAGTAATTTAAATGTAAAATGTATAATGTATAATGTAAAATGTTTTGAAATTGATTTTCAATTAATTTTTCGTGTAACAAAAGAAGAATGAAAAAAAATTTCATTATGGA

1202 ATTTTTGGGGGATATTGAATCGTGTTGGTTAAATACTTTTGTCCTATTCAATTCAGCTTTGAGTTTTACTAGTTTGCCACCTGGTTTGCCACTTAGTTTTGCCACCAAGAAGTTG

1317 GACTAAAGTTTATATCTGTCTCTTATATAATTTACCTTATAGTCAACTTCACTTCCTTCTTTCATTCCTTGTAGTCTTTGTTACATTTTTTGTTCTGGTTCCTGTTACTAACAAC

1432 AACACAATATTTTTTTTTTTTAATTCCCTCACTCAATTCAACTCACAAACCCTATTTCTTTTCTTCTTCTACTTCTTGGTGTAACTTAAACCTTTTTGCTGGCCTCTTTCTTTTG

1547 CTATTTTCTAAAATTCAATTTTGTTAGCTCATTTTAATTATTTCAATTCAATTGTTTTCCATTTATTTCCATCTTTCGTTTTCCATTTATTTTATTTTCTTCTTTTTAGTTAGCT

1662 CTAATTCAACTTCTTCTACTTCTAACTTGAAATCTAACAACTAAAAAAAGACAAGGAGTGAAAAAAAGTGTGAGAATTTGCTAAAAAAAAATAGAGACAGAAAAAGAAAAAGTTA

1777 ACGAACACCAAAGACAGGAAGAAAAAAAAAATTCCAACAACAGGAACAAACATCAACAAACATTAACATCAGCAACAAGCAGAAGACCAATATACATTAACCAATCACACTGAAC

1892 TTACTCATAAACTACTTGCTCATATCTTCCTCTTTTTTTTTTTTTGCTGCATATTGAAGAAATAGAAACCAAATAGAACCACTCATTATATCTTAATATCAACAAATCCAAAACC

Met Glu His Pro Pro Ala Ala Leu Arg Thr Phe Ser Thr Gln Ser Thr Ser Ser Leu Asn Ser Val Ser Thr Val Ser Ser Ser Arg 29

2007 ATG GAA CAT CCA CCA GCA GCT CTC AGA ACA TTT TCA ACC CAA TCA ACT TCA TCT TTG AAT TCA GTA AGT ACT GTT TCG TCT TCA AGA

Ile Val Ser Leu Gly Pro Val Asn Ile Asn Asn Phe Asn Lys Pro Ser Thr Pro Lys Asp His Leu Phe Tyr Arg Cys Glu Ser Leu 58

2094 ATT GTT TCT CTG GGC CCA GTC AAT ATA AAC AAT TTC AAT AAA CCA AGT ACT CCC AAA GAC CAT TTA TTC TAT CGA TGT GAA TCA CTA

Lys Arg Lys Leu Gln Lys Ile Pro Gly Met Glu Pro Phe Leu Asn Gln Ala Phe Asn Gln Ala Glu Gln Leu Ser Glu Gln Gln Ala 87

2181 AAA CGA AAA CTA CAA AAA ATC CCT GGC ATG GAA CCA TTT TTG AAC CAA GCT TTC AAT CAG GCT GAA CAA CTC AGT GAA CAA CAA GCA

Leu Ala Leu Ala Gln Glu Arg Ser Asn Gly Asn Gly His Ser Asn Gly Lys Arg His Gln Ser Leu Asp Gly Ala Met Asn Arg Leu 116

2268 TTG GCT TTG GCA CAG GAA AGA AGC AAT GGA AAT GGA CAT AGT AAT GGC AAA CGT CAT CAA TCA TTA GAC GGT GCC ATG AAT AGA CTT

Ser Val Gly Ser Asp Ser Ser Ser Ile Gln Gly Ser Leu Thr Arg Met Ala Thr Asn Ala Ser Thr Ser Ser Leu Ile Ser Gly Met 145

2355 TCA GTT GGT TCT GAT AGT AGT TCG ATT CAA GGT TCA TTG ACA CGA ATG GCT ACC AAT GCG TCA ACG TCA TCT TTA ATC AGT GGT ATG

FIG. 3A

```
     Pro Asn Ser Asn Thr Leu Phe Thr Phe Thr Ala Gly Val Leu Pro Ala Asn Ile Ser Val Asp Pro Ala Thr His Leu Trp Lys Leu   174
2442 CCA AAC AGC AAC ACT TTA TTT ACG TTT ACT GCA GGG GTT TTA CCA GCT AAT ATT AGT GTC GAT CCT GCT ACC CAT CTT TGG AAA TTG
     Phe Gln Gln Gly Ala Pro Phe Cys Val Leu Ile Asn His Ile Leu Pro Asp Ser Gln Ile Pro Val Val Ser Ser Asp Asp Leu Arg   203
2529 TTC CAA CAA GGG GCC CCC TTT TGT GTT CTT ATC AAT CAT ATC CTT CCT GAT TCC CAA ATA CCA GTT GTC AGT TCT GAT GAC TTG AGA
     Ile Cys Lys Lys Ser Val Tyr Asp Phe Leu Ile Ala Val Lys Thr Gln Leu Asn Phe Asp Asp Glu Asn Met Phe Thr Ile Ser Asn   232
2616 ATT TGC AAA AAA TCA GTA TAT GAC TTT TTA ATT GCC GTC AAG ACG CAA TTG AAT TTT GAT GAC GAG AAT ATG TTC ACT ATA TCC AAT
     Val Phe Ser Asp Asn Ala Gln Asp Leu Ile Lys Ile Ile Asp Val Ile Asn Lys Leu Leu Ala Glu Tyr Ser Asp Ala Ser Asp Leu   261
2703 GTT TTC TCC GAC AAT GCC CAA GAT TTA ATC AAG ATT ATT GAT GTC ATT AAT AAA CTA CTT GCT GAG TAC TCA GAT GCT AGT GAC CTG
     Gly Gly Gly Asp Glu Asp Val Asn Met Asp Val Gln Ile Thr Asp Glu Arg Ser Lys Val Phe Arg Glu Ile Ile Glu Thr Glu Arg   290
2790 GGT GGT GGC GAT GAA GAT GTA AAT ATG GAT GTT CAA ATT ACC GAT GAA AGA TCA AAA GTT TTC CGA GAA ATT ATC GAA ACA GAA AGA
     Lys Tyr Val Gln Asp Leu Glu Leu Met Cys Lys Tyr Arg Gln Asp Leu Ile Glu Ala Glu Asn Leu Ser Ser Glu Gln Ile His Leu   319
2877 AAA TAT GTT CAA GAC TTG GAA CTA ATG TGT AAA TAC CGT CAA GAT CTA ATT GAA GCC GAA AAT TTG TCT TCA GAA CAA ATT CAC TTG
     Leu Phe Pro Asn Leu Asn Glu Ile Ile Asp Phe Gln Arg Arg Phe Leu Asn Gly Leu Glu Cys Asn Ile Asn Val Pro Ile Arg Tyr   348
2964 TTA TTC CCA AAT TTA AAT GAG ATT ATT GAT TTT CAA AGA CGA TTC CTC AAT GGG TTA GAA TGT AAC ATC AAT GTA CCT ATT AGA TAT
     Gln Arg Ile Gly Ser Val Phe Ile His Ala Ser Leu Gly Pro Phe Asn Ala Tyr Glu Pro Trp Thr Ile Gly Gln Leu Thr Ala Ile   377
3051 CAA AGA ATT GGA TCA GTA TTT ATT CAT GCT TCT TTG GGC CCT TTC AAT GCT TAT GAA CCT TGG ACT ATA GGA CAA TTG ACG GCG ATT
     Asp Leu Ile Asn Lys Glu Ala Ala Asn Leu Lys Lys Ser Ser Ser Leu Leu Asp Pro Gly Phe Glu Leu Gln Ser Tyr Ile Leu Lys   406
3138 GAT TTG ATC AAC AAA GAA GCT GCT AAT TTG AAA AAA TCA TCG AGT CTA CTT GAT CCT GGG TTT GAA CTT CAA TCG TAT ATA TTA AAG
     Pro Ile Gln Arg Leu Cys Lys Tyr Pro Leu Leu Leu Lys Glu Leu Ile Lys Thr Ser Pro Glu Tyr Ser Lys Gln Asp Pro His Gly   435
3225 CCG ATC CAA AGA TTG TGT AAA TAC CCA CTT TTG TTG AAA GAG TTA ATC AAA ACA TCA CCA GAA TAT TCA AAA CAG GAC CCC CAT GGC
     Ser Ser Ser Leu Thr Ser Phe Asn Glu Leu Leu Val Ala Lys Thr Ala Met Lys Glu Leu Ala Asn Gln Val Asn Glu Ala Gln Arg   464
3312 AGC TCG TCA TTG ACA TCA TTC AAT GAA TTA TTG GTG GCT AAA ACT GCA ATG AAA GAA TTG GCA AAT CAA GTC AAT GAG GCG CAA AGA
     Arg Ala Glu Asn Ile Glu His Leu Glu Lys Leu Lys Glu Arg Val Gly Asn Trp Arg Gly Phe Asn Leu Asp Ala Gln Gly Glu Leu   493
3399 CGA GCA GAA AAT ATC GAA CAT TTG GAA AAA CTA AAA GAA AGA GTA GGT AAT TGG CGT GGG TTT AAT TTG GAT GCT CAA GGA GAA CTA
     Leu Phe His Gly Gln Val Gly Val Lys Asp Ala Glu Asn Glu Lys Glu Tyr Val Ala Tyr Leu Phe Glu Lys Ile Val Phe Phe Phe   522
3486 TTA TTC CAC GGA CAA GTT GGG GTT AAA GAT GCT GAA AAT GAA AAG GAA TAC GTT GCT TAT CTT TTT GAA AAA ATT GTA TTT TTT TTC
     Thr Glu Ile Asp Asp Asn Lys Lys Ser Asp Lys Gln Glu Lys Lys Ser Lys Phe Ser Thr Arg Lys Arg Ser Thr Ser Ser Asn Leu   551
3573 ACA GAA ATT GAT GAT AAC AAA AAA TCT GAT AAA CAG GAA AAG AAG AGC AAG TTT TCG ACA AGA AAG AGA TCA ACT TCA TCA AAT CTT
```

FIG. 3A (cont.)

Ser Ser Ser Thr Thr Asn Leu Leu Glu Ser Ile Asn Asn Ser Arg Lys Asp Asn Thr Leu Pro Leu Glu Leu Lys Gly Arg Val Tyr 580
3660 AGT TCA TCG ACT ACT AAT TTG TTG GAA TCA ATA AAC AAT TCC CGA AAG GAT AAC ACA TTG CCA TTG GAA TTA AAA GGA AGA GTT TAT
Ile Ser Glu Ile Tyr Asn Ile Ser Ala Pro Asn Thr Pro Gly Ser Thr Leu Ile Ile Ser Trp Ser Gly Arg Lys Glu Ser Gly Ser 609
3747 ATA TCG GAG ATT TAT AAC ATT TCC GCT CCA AAC ACT CCT GGC TCA ACC CTA ATC ATC TCA TGG TCA GGT AGA AAG GAA AGC GGC TCA
Phe Thr Leu Arg Tyr Arg Ser Glu Glu Ala Arg Asn Gln Trp Glu Lys Cys Leu Arg Asp Leu Lys Thr Asn Glu Met Asn Lys Gln 638
3834 TTC ACT TTG AGA TAT CGT AGT GAA GAA GCC AGA AAC CAA TGG GAA AAG TGT TTA CGT GAT TTG AAG ACT AAT GAA ATG AAT AAA CAA
Ile His Lys Lys Leu Arg Asp Ser Asp Leu Ser Phe Asn Thr Asp Asp Ser Ala Ile Tyr Asp Tyr Thr Gly Ile Ser Thr Ser Pro 667
3921 ATT CAT AAG AAG TTA CGT GAT TCC GAC CTG TCA TTT AAT ACT GAT GAC TCT GCC ATA TAT GAT TAC ACG GGT ATT AGT ACG TCA CCA
Val Asn Gln Ser Thr Gln Gln Gln Tyr Tyr Asp His Arg Gly Ser His Ser Ser Arg His His Ser Ser Ser Ser Thr Leu Ser Met 696
4008 GTC AAT CAA TCA ACT CAA CAA CAA TAC TAT GAT CAT CGG GGC TCT CAC AGT TCC CGC CAT CAC TCA TCG TCA TCC ACT TTG AGT ATG
Met Lys Asn Asn Arg Val Lys Ser Gly Asp Leu Ser Arg Ile Ser Ser Thr Ser Thr Thr Leu Asp Ser Phe Ser Asn Asn Leu Asn 725
4095 ATG AAG AAT AAT AGA GTT AAA TCT GGT GAT TTG AGT AGA ATA TCT TCA ACT TCA ACA ACA TTA GAT TCT TTC AGT AAC AAC TTG AAT
Gly Ser Pro Asn Thr Thr Asn Pro Ser Leu Thr Ser Ser Asp Ala Thr Lys Thr Ile Pro Thr Phe Asp Val Ala Ile Lys Leu Leu 754
4182 GGG TCA CCA AAT ACC ACT AAT CCA TCT TTG ACG TCT TCA GAT GCC ACC AAA ACA ATT CCA ACA TTT GAC GTT GCA ATT AAA TTG CTT
Tyr Lys Ser Thr Glu Leu Ser Glu Pro Leu Ile Val Asn Ala Gln Ile Glu Tyr Asn Asp Leu Leu Gln Lys Ile Ile Ser Gln Ile 783
4269 TAC AAA TCG ACA GAA TTG TCA GAG CCA TTG ATT GTC AAT GCA CAA ATT GAG TAT AAT GAC CTT TTA CAG AAA ATT ATC TCC CAG ATT
Ile Thr Ser Asn Leu Val Ala Asp Asp Val Asn Ile Ser Arg Leu Arg Tyr Lys Asp Asp Glu Gly Asp Phe Val Asn Leu Asn Ser 812
4356 ATC ACT TCG AAC TTG GTG GCT GAT GAT GTC AAT ATT AGT CGA TTG AGA TAT AAA GAC GAC GAA GGA GAC TTT GTG AAT TTG AAT TCA
Asp Asp Asp Trp Gly Leu Val Leu Asp Met Leu Thr Ser Glu Asp Phe Tyr Gln Thr Ser Ser Asn Glu Lys Arg Leu Val Thr Val 841
4443 GAT GAT GAT TGG GGG TTA GTG CTT GAT ATG TTA ACC AGT GAA GAC TTT TAC CAA ACA TCA AGC AAT GAA AAA CGA CTG GTG ACA GTG
Trp Val Ser Stop 844
4530 TGG GTT TCT TGA TTTAACTACAGGAACAAACGCTACCTTTGTTTGGTGTGTGTGTATGTATGGGTGCTTTTTTTTTTTATTTCTTGATGGTGTGTGACTTTGGAAGATA
4641 AACAAATTAAGAGTTAATGTTTTGCTGTGCAAAATAAGCTGTTATAGATGGGTTCAATTAATCAATTTCATATAGATATATAAATGACACTTTGACGAAATATACTATTTATAAA
4756 TTTCCTTTTTCTTTGTTTTGTAAGATTAATGTTGGTTCTTGTTGATGTGTCGGTACACCAAACGCAATAATTAAAATCTAGTAAGACGGTAAATGGGTAGATGAGAAAAGGTCA
4871 ATAGAGTTTATTCTAATGTGGGTGCAAATTAAAGGCAACAGATAAATTTGGTAAACATTTTCTAAAACGTATTGCCGCTTCCAGAGTCAAAAAAAAGAATAAAGCTAATATATTA
4986 GTGCTAATAATAGTAGTAATACAAAACAAGGTTTCAAAGTTTTCGCTCAAAACATCAAGCCATTGCTTATATAGGATGAACTATTCAATTAACAGGCAAAAAAAAGCCATCATTT
5101 GAAAAGACTCTCATATCAAAGAGGTAACTTCTAATAGTAATCACTTGTTGTTTTTGATTATTAAATGATTTGATTCTATTGGTTGAACTAACCCCAAWTGGGTTTKTTGTTTGCC
5216 GGGTTGARAATGAATGCCATAAATNATTCAATTTGAAAAAAAAAAAAAATNCTAATACAACACACCCAACCCTTTGCNTTTATCA

FIG. 3A (cont.)

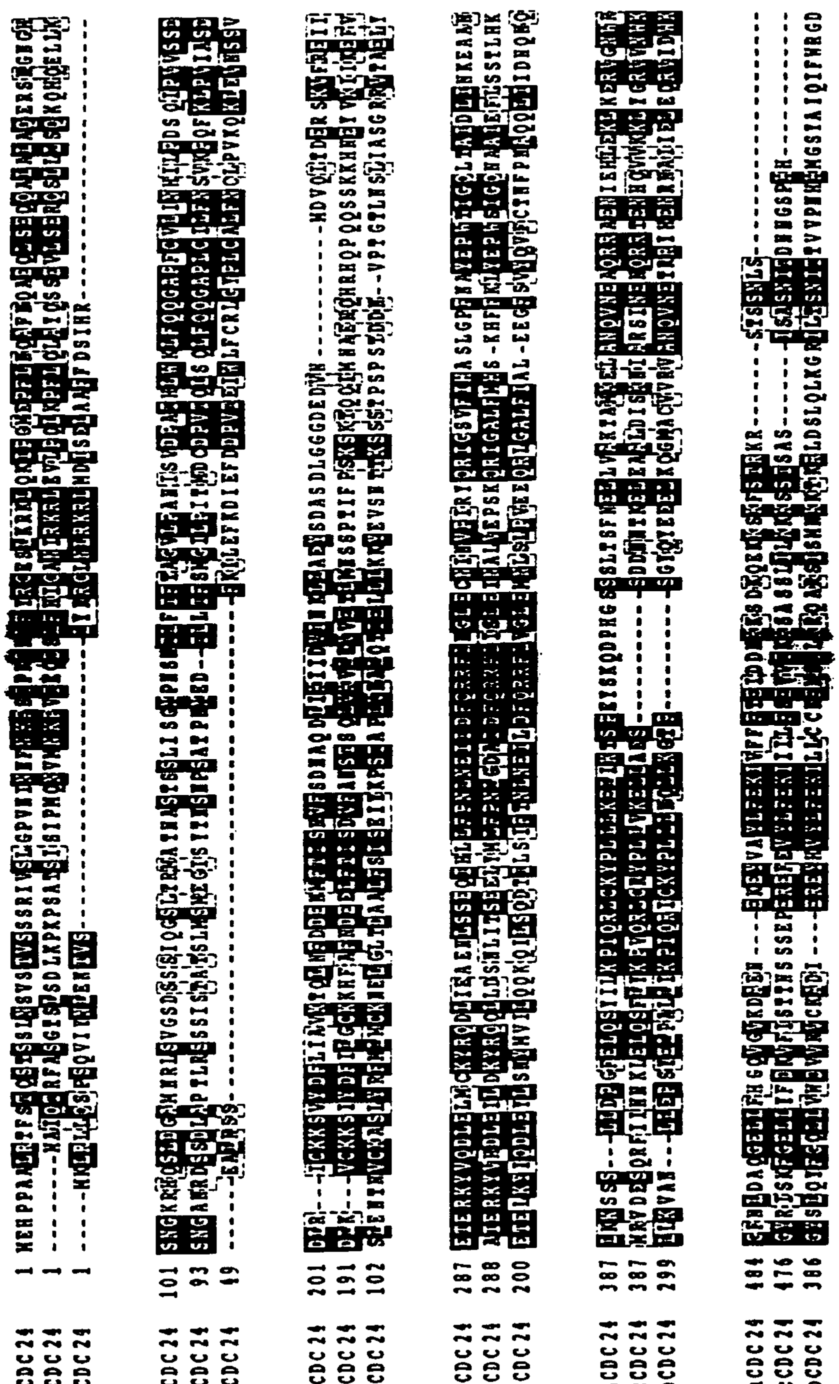

FIG. 3B

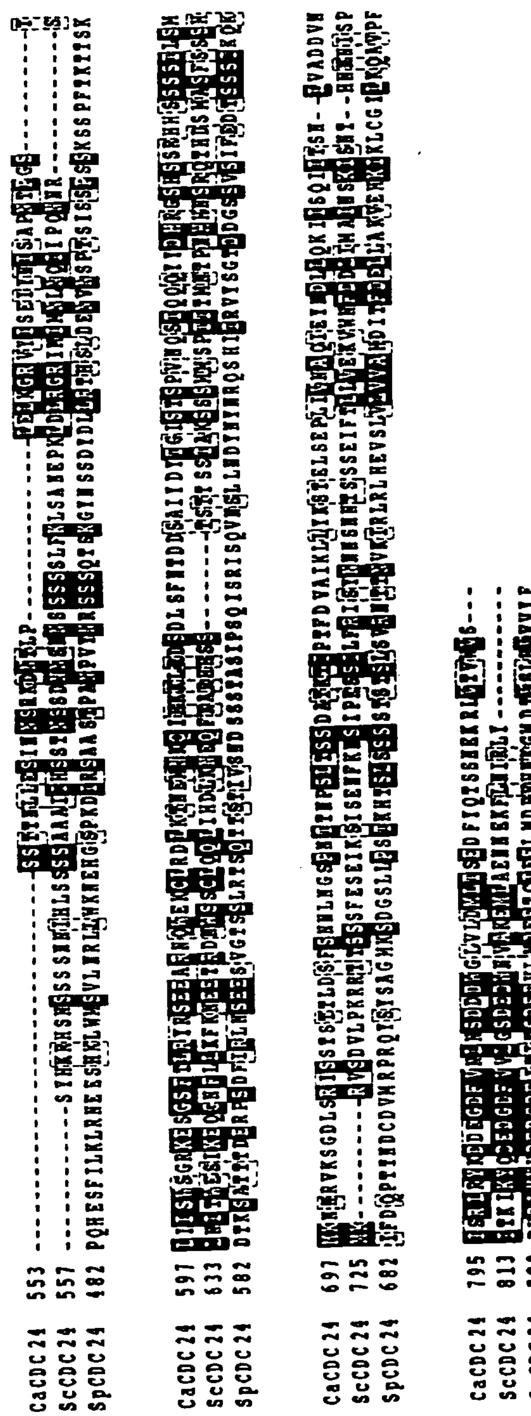

FIG - 3B (cont.)

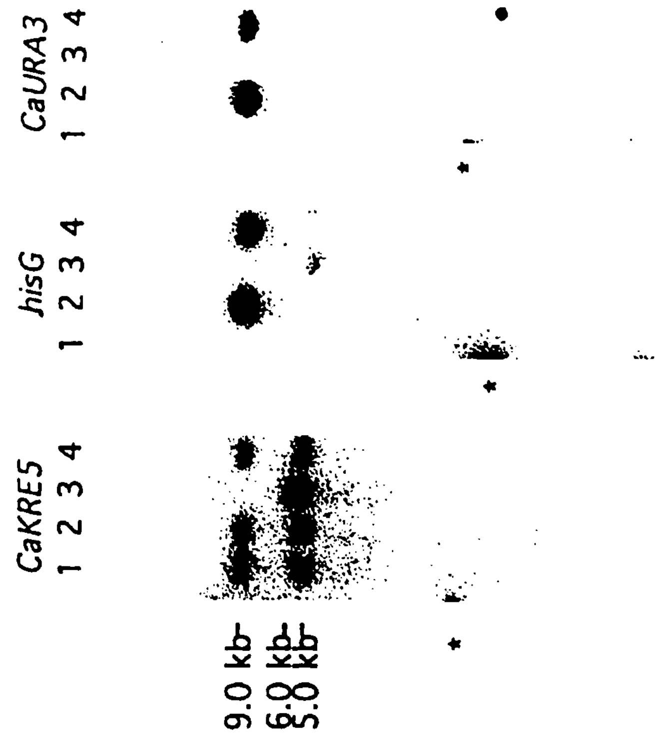
CaKRE5
1 2 3 4
hisG
1 2 3 4
CaURA3
1 2 3 4
9.0 kb-
6.8 kb-
5.0 kb-

Fig. 4B

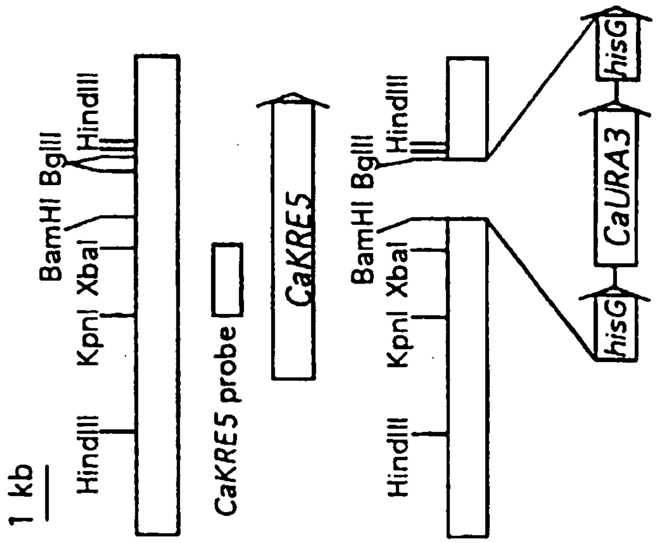
1 kb
HindIII
KpnI
XbaI
BamHI
BglII
HindIII
CaKRE5 probe
CaKRE5
HindIII
KpnI
XbaI
BamHI
BglII
HindIII
hisG
CaURA3
hisG

Fig. 4A

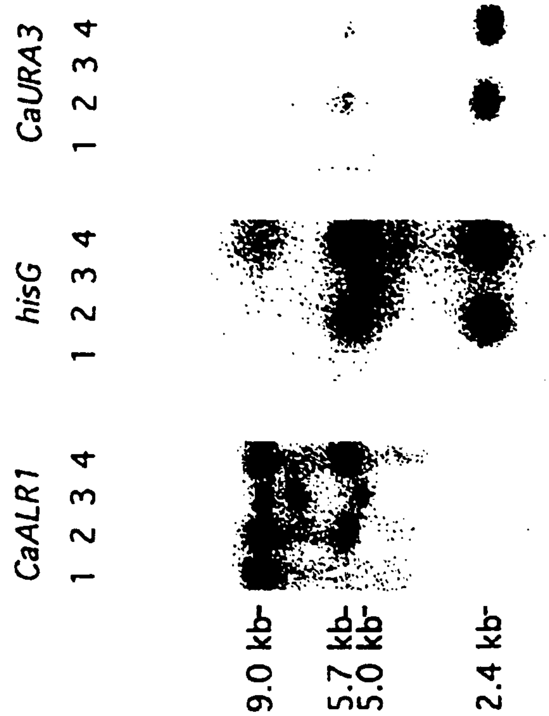
CaALR1
1 2 3 4
hisG
1 2 3 4
CaURA3
1 2 3 4
9.0 kb
5.7 kb
5.0 kb
2.4 kb

FIG. 4D

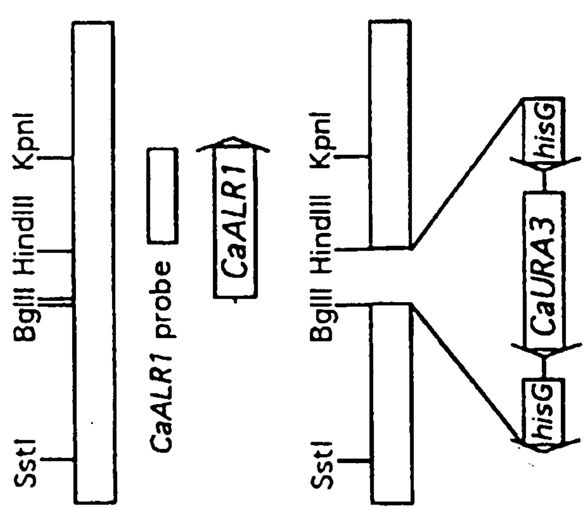
SstI
BglII
HindIII
KpnI
CaALR1 probe
CaALR1
SstI
BglII
HindIII
KpnI
hisG
CaURA3
hisG

FIG. 4C

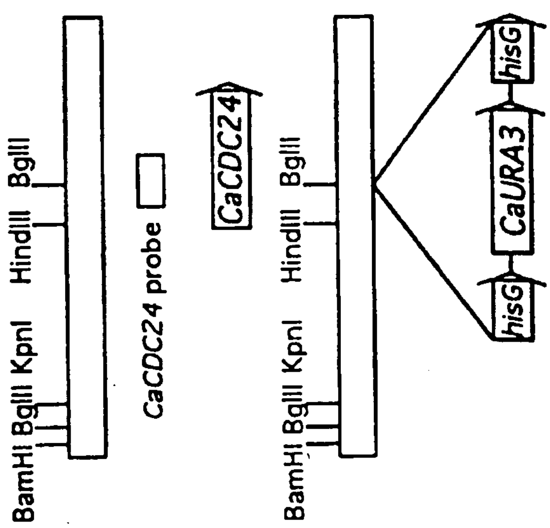
BamHI BglII KpnI HindIII BglII
CaCDC24 probe
CaCDC24
BamHI BglII KpnI HindIII BglII
hisG
CaURA3
hisG

Fig. 4E

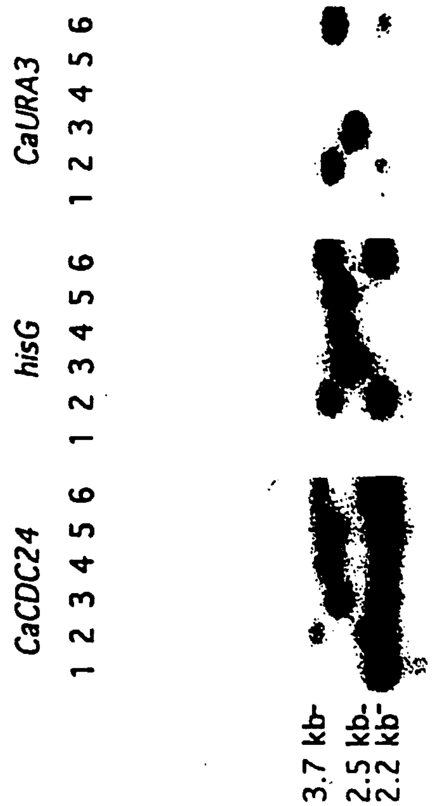
CaCDC24
1 2 3 4 5 6
hisG
1 2 3 4 5 6
CaURA3
1 2 3 4 5 6
3.7 kb
2.5 kb
2.2 kb

Fig. 4F